Grapevine field experiments reveal the contribution of genotype, the influence of environment and the effect of their interaction (GxE) on berry

Article type : Resource

Grapevine field experiments reveal the contribution of genotype, the influence of environment and the effect of their interaction (GxE) on berry transcriptome

Silvia Dal Santo*§[1], Sara Zenoni§[1], Marco Sandri[1], Gabriella De Lorenzis[2], Gabriele Magris[3], Emanuele De Paoli[4], Gabriele Di Gaspero[3], Cristian Del Fabbro[4], Michele Morgante[3], Lucio Brancadoro[2], Daniele Grossi[2], Marianna Fasoli[1], Paola Zuccolotto[5], Giovanni Battista Tornielli[1] and Mario Pezzotti[1]*

[1] Department of Biotechnology, University of Verona, I-37034, Verona, Italy

[2] Department of Agricultural and Environmental Sciences - Production, Landscape, Agroenergy, University of Milano, I-20133, Milano, Italy

[3] I.G.A., Applied Genomics Institute, I-33100, Udine, Italy

[4] Department of Agricultural, Food, Environmental and Animal Sciences (DI4A), University of Udine, I-33100, Udine, Italy

[5] Department of Economics and management, University of Brescia, I-25121, Brescia, Italy

§ Equal contribution

* Corresponding authors: Silvia Dal Santo (silvia.dalsanto@univr.it); Mario Pezzotti (mario.pezzotti@univr.it)

Running head: GxE Interaction molecular dissection in grapevine

KEYWORDS: Genotype x Environment Interaction (GxE); *Vitis vinifera* (grapevine), Data mining, Gene Expression variation, Secondary Metabolism;

SUMMARY

Changes in the performance of genotypes in different environments are defined as genotype x environment (GxE) interactions. In grapevine (*Vitis vinifera*), complex interactions between different genotypes and climate, soil, and farming practices yield unique berry qualities. However, the molecular basis of this phenomenon remains unclear. To dissect the basis of grapevine GxE interactions, we characterized berry transcriptome plasticity, genome methylation landscape, and within-genotype allelic diversity in two genotypes, cultivated in three different environments, over two vintages. We identified, through a novel data-mining pipeline, genes with expression profiles that were unaffected by genotype or environment, genotype-dependent but unaffected by the environment, environmentally-dependent regardless of genotype, and GxE-related. The GxE-related genes showed different degrees of within-cultivar allelic diversity in the two genotypes and were enriched for stress responses, signal transduction and secondary metabolism categories. Our study unraveled the mutual relationships between genotypic and environmental variables during GxE interaction in a woody perennial species, providing a reference model to explore how cultivated fruit crops respond to diverse environments. Also, the pivotal role of vineyard location in determining the performance of different varieties, by enhancing berry quality traits, was unraveled.

INTRODUCTION

The phenotype of every organism is determined by a combination of its genotype (G), environment (E) and genotype-dependent responses to different environments, the latter known as genotype x environment (GxE) interactions (Grishkevich and Yanai, 2013, El-Soda *et al.*, 2014). Variations in gene expression reflecting different types of genetic and epigenetic regulation can be used as a proxy to define genotype–phenotype relationships in a changing environment (Rockman and Kruglyak, 2006, Perry and Mank, 2014). Recent developments in genomics and genome-wide transcriptome profiling have therefore revolutionized molecular ecology and evolutionary genetics, offering opportunities to expand traditional GxE studies beyond model organisms (Thomas, 2010, Perry and Mank, 2014).

Plants have a remarkable ability to thrive despite their limited capacity to alter their surroundings (Des Marais *et al.*, 2013). This phenomenon relies on phenotypic plasticity (the ability to express different phenotypes from the same base genotype depending on the circumstances) and has gained attention recently due to the challenges posed by climate change (Nicotra *et al.*, 2010). The stability of crop growth and yields must be maintained over diverse and dynamic environments, and an understanding of how the genotype responds to and

interacts with the environment is necessary to predict the effects of climate change on ecology and modern agriculture (Fournier-Level *et al.*, 2011, Sasaki *et al.*, 2015). However, the environmental component of this complex interaction is often expensive or impossible to define with any precision in natural environments, and studies based on gene expression variation in open-field-grown plants do not tend to address GxE interactions in detail (Brosché *et al.*, 2005, Holliday *et al.*, 2010, Travers *et al.*, 2010, Richards *et al.*, 2012, Dal Santo *et al.*, 2013, Dal Santo *et al.*, 2016b, Hess *et al.*, 2016).

Grapevine (*Vitis* spp., family Vitaceae) is an economically important fruit crop used globally to produce food and beverages. This crop is characterize by a pronounced sensitivity towards the environment, and the metabolic composition of the berries is characterized by broad phenotypic plasticity, offering advantages such as the range of different wines that can be produced from the same cultivar, and the adaptation of existing cultivars to different growing regions (Keller, 2010, Dai *et al.*, 2011). The relevance of the interaction between varietal genotypes and the environment is best exemplified by the concept of *terroir*, which combines varietal attributes with the climate, soil and winemaking practices, plus all the possible interactions among them. It is anecdotally known that many grapevine varieties perform differently in distinct environments, with some varieties such as Cabernet Sauvignon and Chardonnay offering more consistency, and others such as Sangiovese, Nebbiolo and Pinot noir showing greater variation. Most grapevine GxE studies have focused on single traits using classical methods such as the analysis of quantitative trait loci (Adam-Blondon *et al.*, 2011), but we have recently explored the use of "omics" approaches to unravel the phenotypic plasticity of grapevine berries on a broader scale (Dal Santo *et al.*, 2013, Anesi *et al.*, 2015, Dal Santo *et al.*, 2016b, Paim Pinto *et al.*, 2016).

Here we investigated the phenotypic plasticity and GxE interactions of two grapevine varieties by analyzing their transcriptomes in three different environments at four different developmental stages over two consecutive vintages. A tailored statistical data-mining tool based on data reduction allowed the inspection of G, E and GxE clusters of gene expression, and contributed to the identification of several candidate genes that could be used as markers of berry quality traits in GxE interactions. Parallel genomic and epigenomic analysis provided a multi-layered scientific definition of the formerly empirical basis of *terroir*. Finally, correlation analysis was applied to the transcriptomic and climatic data to unravel the molecular basis of GxE interactions in open-field-grown crops.

RESULTS

Experimental design of the GxE interaction studies

Grapevine berries (*V. vinifera* cultivars Sangiovese and Cabernet Sauvignon) were harvested at four different developmental stages – pea size (PS), pre-veraison (PV), mid-ripening (MR) and fully ripe (FR) – from three central Italian locations (Bolgheri on the Tuscany coast, Montalcino in the Tuscany hills, and Riccione on the Adriatic coast) during the 2011 and 2012 growing seasons (Figure 1a; Tables S1 and S2). The berries were collected in biological triplicates, giving a total of 144 samples (Table S3). We recorded the daily mean temperature (Td), daily maximum temperature (Tx), global solar radiation (GSR), growing degree days (GDD), rainfall, and available soil water content (AWC) throughout the experiment (Figure 1a; Figure S1). Climatic parameters differed among the locations and vintages, with the largest differences recorded in Bolgheri for the lower Td values and in Montalcino for the highest GSR. Interestingly, AWC data revealed water stress in all three vineyards, between June and September 2011 and between June and August 2012 (Figure 1a; Text S1).

Fruit composition and yield components were evaluated at harvest in the 2011 and 2012 seasons. There were statistically significant differences in each of the parameters, except for the Sangiovese yield per vine and number of berry clusters in 2012 (Table S4). In particular, the highest soluble solids content in both varieties (°Brix) was recorded in the Riccione (2011) and Montalcino (2012) regions (Figure 1b, inset). The lowest berry weights at all developmental stages were recorded in the Montalcino region, with the exception of the most variable PS stage (Figure 1b).

The physiological response of the vines to environmental variables was assessed by monitoring trends in the photosynthetic rate (Pn), stomatal conductance (gs), transpiration rate (E_T), soil water content (SWC), and stem water potential (SWP). This analysis revealed that the Montalcino region suffered the greatest degree of water stress during both growing seasons (Text S1). We also monitored the carotenoid, norisoprenoid, chlorophyll, flavonol and hydroxycinnamic acid (HCA) content of the berries (Tables S5 and S6), revealing a general positive relations for both varieties during early developmental stages before veraison between carotenoid levels and the regional GSR, which was highest in Montalcino and Bolgheri (Figure 1A). The synthesis of norisoprenoid compounds in Sangiovese berries varied among the locations and vintages, and appeared more dependent on ecophysiological conditions during maturation than the carotenoid content (Text S1).

Sangiovese berries show greater transcriptomic plasticity than Cabernet Sauvignon

The plasticity of the grapevine berry transcriptome in response to environmental variables was determined using the NimbleGen whole-genome microarray (090918_Vitus_exp_HX12). A Pearson's distance correlation matrix was generated to compare the 48 berry transcriptomes (Figure 2a) revealing a strong correlation (R>0.85) between samples collected before the onset of ripening (PS and PV), and between samples collected during ripening (MR and FR), regardless of cultivar, vintage and location, as previously reported for Corvina berries (Fasoli *et al.*, 2012). The correlation values were used as distance coefficients to build a dendrogram, which described the dynamic berry transcriptome in greater depth (Figure 2b). The pre-ripening samples clustered largely according to the maturation stage whereas the vineyard location had no significant impact. Similarly, the post-ripening Cabernet Sauvignon samples revealed a stable clustering pattern based on stage>vintage>location, but in the Sangiovese samples this hierarchy was only observed for the FR berries collected in 2012 (Figure 2b). The number of transcripts showing significant modulation between vintages and among locations was assessed separately in the two genotypes, firstly by overcoming the typical bi-modal distribution of Nimblegen-derived fluorescence intensity values (Figure S2), then by two-way analysis of variance (ANOVA). This analysis revealed that ~25% of the modulated genes in each genotype were differentially expressed between the 2011 and 2012 vintages (Figure 2c), agreeing with previous reports showing the impact of vintage on berry transcriptome plasticity (Dal Santo *et al.*, 2013). However, the effect of location was greater in Sangiovese than Cabernet Sauvignon, with almost twice as many genes in the former cultivar differentially expressed among the three locations as well as in the vintage x location interaction (Figure 2c), indicating a greater degree of transcriptomic plasticity in Sangiovese berries, under our experimental conditions,.

The potential epigenetic basis of these cultivar-dependent differences was investigated by comparing the DNA methylation level in the PV and MR samples (two cultivars, three locations, two vintages) by reduced representation bisulfite sequencing. All samples provided comparable methylation data for a subset of ~23,000 cytosine residues enriched in the genic compartment, particularly at the 5' end of transcribed regions (Figure S3a). The genotype appeared to be a major covariate accounting for up to 39% of methylation variance between samples, depending on the sequence context (Figure 2d-f, Figure S3b-d), and was associated with significant differences in methylation across the cytosine panel (Figure S3e-h). Significant hypermethylation was consistently observed at CHH sites during the MR developmental stage (Figure S3i). However, there was no convincing association between methylation and environmental conditions, indicating that methylation remained stable regardless of variations in external cues and in gene expression.

Grapevine GxE interactions revealed by a novel statistical approach

The large scale of our sampling procedure required the development of a new statistical approach to uncover the hidden GxE interactions and to determine how they affect berry transcriptome plasticity in field-grown plants. A three-step data mining pipeline (Figure 3a, Text S2) was therefore used to summarize the most important relationships within the dataset, focusing on the quantitative impact of stage, cultivar, vintage and location (and interactions among them) on gene expression.

***Step 1: Screening*.** We identified a subset of 11,427 genes with uninteresting profiles, i.e. no expression, constitutive expression or outlier expression (Figure S4 and Table S7). The remaining dataset thus comprised 18,122 genes warranting statistical analysis (Dataset S1).

***Step 2: Cluster definition*.** We applied k-means clustering to the subset of 18,122 interesting genes, resolving to 300 clusters that accounted for ~70% of the total variance in gene expression (Figure S5a). For each cluster, we defined an average representative expression profile and an index of its representativeness (homogeneity index, R_c) based on the variability of expression around the average profile, which measured the internal cohesion of each cluster (Figures S5b-c).

***Step 3: Cluster characterization*.** We then used an advanced machine learning algorithm known as the Gradient Boosting Machine (GBM)(Friedman, 2001) to evaluate the extent to which each of the variables (stage, cultivar, vintage and location) affected gene expression. The GBM output was a set of variable importance measures (VIMs), i.e. nonparametric statistical tools that estimate the impact of covariates on a selected outcome, taking into account the effect of potential (even complex) interactions among variables and nonlinear relationships with the outcome. The median VIMs of each of the 300 clusters were used to characterize the relationship between the clusters and the four experimental conditions (Text S2). Principal component analysis (PCA) was then used to reduce the dimensionality of the resulting matrix, in which the average profiles of the 300 clusters were arranged as columns. Principal components, computed as linear combinations of cluster profiles, were able to discriminate among the stage, cultivar and vintage variables characterizing the 48 experimental conditions with a remarkable accuracy (Figure S6). Figures 3b-e show that the loadings of the clusters in the first, second, third and tenth rotated principal components (DimRot1, 2, 3 and 10) are associated with the importance of the stage, cultivar, vintage and location variables, respectively. The location variable showed the weakest association of loadings and least importance, and homogeneity within these clusters was low. The location-related clusters also presented more complex profiles, which appeared to be affected by interactions with other variables (Figure 3e).

In summary, the new statistical pipeline allowed the 18,122 modulated genes to be assigned to 300 clusters, each described by four VIMs (one for each variable). Each VIM has its own dynamic range due to the intrinsic importance of that variable in explaining the total variability of the dataset, resulting in the maximum dynamic range for the stage variable and the minimum range for the location variable. We therefore assigned a rank to each cluster according to the VIM for each variable. For example, cluster #266 has similar values for VIM_Location (196.46) and VIM_Stage (177.70) and is ranked first for the location variable but only 282nd for the stage variable (Dataset S2 and Dataset S3).

Influence of variables on transcriptional variation in the context of GxE interactions

A rank-based approach was developed to classify the clusters. Variable-specific clusters were defined as those ranking in the top 100 for only one of the four variables, whereas variable-shared clusters were defined as those ranking in the top 100 clusters for more than one variable (Dataset S2). The specific and shared clusters were mapped using a Venn diagram (Figure 4a).

Most of the clusters (75) were stage-specific, comprising 6,793 genes and accounting for 37.5% of all modulated genes (Figure 4b, Dataset S2 and Dataset S4). BINGO Gene Ontology (GO) enrichment analysis applied to genes in the 75 stage-specific clusters revealed enriched functional categories related to photosynthesis and energy generation, response to endogenous stimuli, and carbohydrate metabolism (Figure 4c; Text S3). Interestingly, the number of stage-specific clusters with a downregulated metaprofile (38, comprising 3,243 genes) was nearly identical to the number showing upregulation during berry ripening (37, comprising 3,550 genes) (Figure 4d). Stage-specific transcripts were transcribed from genes located predominantly in distal chromosome regions, whereas pericentromeric genes were significantly underrepresented, with 197 cases compared to 329.1–331.1 expected within the confidence interval (Figure 4e, Dataset S5).

There were 48 cultivar-specific clusters, containing 2,648 genes and accounting for 14.6% of all modulated genes (Figure 4b, Dataset S2 and Dataset S4). These were mainly enriched for functional categories related to biotic and abiotic stress, such as response to stress, death, and cell death (Figure 4f; Text S3). An analysis of copy number variation (CNV) identified 52 differentially expressed genes in genomic regions differing in copy number between the Cabernet Sauvignon and Sangiovese cultivars. Cluster analysis classified 39 of these transcripts as cultivar-specific, and in 31 cases the difference in copy number was concordant with the difference in absolute transcript levels determined by RNA-seq analysis (Figure 4g, Dataset S6). The remaining cultivar-specific transcripts were also transcribed from genes that varied in copy number between the cultivars, but the cultivar with fewer copies showed higher expression

levels. However, in all these cases the genes were minimally expressed in both cultivars based on a mean fragments per kilobase mapped (FPKM) value of less than 1 (Dataset S6).

There were 26 vintage-dependent clusters, containing 1,657 genes and representing 9.1% of all modulated genes (Figure 4b, Dataset S4). These were enriched for cellular process and signal transduction functions (Figure 4h; Text S3) and contained many signal transduction effectors, including components of calcium-based signaling pathways (calmodulins, calcium-binding proteins and calcium-dependent protein kinases). These are used in a flexible manner by plants to couple variable external signals to specific cellular responses(Yang and Poovaiah, 2003).

Finally, there were 27 location-specific clusters, containing 1,183 genes and representing 6.5% of all modulated genes (Figure 4b, Dataset S4). These clusters were characterized by a smaller average number of genes per cluster and a lower average Rc index than the other variable-specific clusters. Only 12 of the clusters (44%) ranked among the top 50 $VIM_{Location}$ scores (Dataset S2), indicating that the location *per se* contributes less to variations in berry gene expression than the other variables. However, the 27 location-specific clusters were particularly enriched for the functional category secondary metabolic process (Figure 4i). For example, they included several members of the stilbene synthase gene family, which control resveratrol synthesis, as well as genes responsible for monoterpene synthesis and the oxidative polymerization of phenolic compounds in the phenylpropanoid pathway (Pourcel *et al.*, 2005) (Text S3).

As stated above, clusters in the top 100 of more than one VIM ranking were defined variable-shared clusters. We identified 106 variable-shared clusters comprising 4,876 genes representing 26.9% of all modulated genes (Figure 4a-b, Dataset S7). The most important association in terms of the number of clusters and genes was observed between the vintage and location variables (39 clusters, 1,478 genes), suggesting that the mutual relationships among different vintages and geographical sites are critical determinants of berry transcriptomic plasticity (Figure S7 and Text S3). The variable-shared clusters associating cultivar and vintage, cultivar and location, or cultivar, vintage and location, represent that part of the grapevine transcriptome specifically involved in GxE interactions (Figure 5, Dataset S7). These associations included 42 clusters and 1,718 genes enriched in the functional categories death, cell death, response to stress, signal transduction, and secondary metabolic process (Figure 5b-e). Interestingly, these GxE clusters also featured genes representing the general phenylpropanoid pathway, lignin biosynthesis, anthocyanin metabolism and the production of volatile metabolites (Text S3).

Next we considered the role of genetic diversity between and within cultivars as a potential explanation for the differences in gene expression profiles in relation to environmental

variables and interactions. Differentially expressed genes were classified based on the level of haplotype sharing between the Cabernet Sauvignon and Sangiovese cultivars. We found that 966 genes were located in 14 Mb of genomic DNA that is fully conserved between the cultivars, whereas 10,094 genes were located in 164.4 Mb in which the two varieties shared one haplotype, and as many as 15,244 genes were located in 240 Mb with no haplotype sharing (Figure 6a, Dataset S8). Cultivar-specific clusters were significantly enriched in transcripts from genes with no haplotype sharing (sharing 0) and depleted in transcripts from genes with haplotype sharing (sharing 1 or 2), whereas stage-specific clusters were significantly enriched in transcripts from genes with partial haplotype sharing (Figure 6b). The role of within-cultivar diversity was considered in more detail by classifying the 18,122 modulated genes according to the zygosity of the corresponding locus in each cultivar (Dataset S9). A chi-squared analysis revealed that loci that are homozygous in Cabernet Sauvignon and heterozygous in Sangiovese, or vice versa, were overrepresented in clusters of transcripts that explain GxE interactions (Figure 6c).

Correlation between transcriptomic and climatic/physiological data unravels the GxE interactions in grapevine

Relationships between the retrieved transcriptomic data and environmental data were determined by Spearman's correlation analysis of the 48 sampling conditions (two cultivars, four stages, three locations and two vintages) in terms of both gene expression (the average gene expression in each of the 300 clusters) and relevant environmental features. Some physiological/biochemical parameters were also included to highlight the phenotype-related effects of GxE interactions. The results are represented by the heat map in Figure 7a (left panel) and the data are shown in Dataset S10. The expression profiles of the 300 clusters showed significant correlation with certain parameters during pre-veraison berry development (e.g. total chlorophyll, carotenoid and organic acid levels, Pn, E_T and gs) and others more relevant during ripening (e.g. total anthocyanin content, berry weight, total GSR, GDD and heat waves index). Clusters showing the highest positive or negative correlations with environmental parameters tended to be those ranked in the first positions for the VIM of the stage variable (Figure 7a, right panel). As expected, clusters correlating strongly with pre-veraison parameters were characterized by downregulated expression (negative DimRot1 parameter, Figure 7a central panel), whereas clusters correlating strongly with post-veraison parameters were characterized by upregulated expression (positive DimRot1 parameter, Figure 7a central panel). In contrast, parameters calculated as mean values during the 5 days before each sampling date (temperature-related parameters and rainfall) showed few high-correlation values with the

expression profiles. Interestingly, the heat map revealed several cases of strong correlation also for clusters highly ranked in the VIMs of the cultivar, vintage and location variables (Figure 7a, left and right panels), indicating that these variables show more hidden but still retrievable relationships with the environmental/biochemical parameters. These results prompted us to repeat the correlation analysis separately for the pre-veraison and post-veraison phases and for the Sangiovese and Cabernet Sauvignon samples, resulting in four correlation matrices containing 12 experimental observations each: one cultivar, two stages, three locations and two vintages (Figure S8, Dataset S11). We then calculated the subtraction matrices for the Sangiovese and Cabernet Sauvignon correlation matrices at each developmental phase. This allowed us to retrieve clusters in which the difference between the two cultivars differed most significantly in terms of the interaction with the environment. The pre-veraison subtraction matrix (Figure 7b, Dataset S12) reveled that temperature, Rainfall 5D and GSR maximize genotype-dependent transcriptomic plasticity, whereas the cultivars become more distinct as maturation proceeded, particularly in terms of the photosynthesis-related parameters (Pn, E_T and gs) and the reaction to rainfall 5D and heat waves (Figure 7c, Dataset S12). For example, in the pre-veraison phase, Cluster #92 (R_c = 0.74), exhibiting a significant negative correlation with Td_5D, Tx_5D, and the HWI index only in Cabernet Sauvignon, encompassed many transcripts for anthocyanins and flavonols metabolism. Also, Cluster #30 (R_c = 0.74), exhibiting a significant negative correlation with the stomatal conductance gs only in Cabernet Sauvignon, contained the *VvNCED1* transcript encoding for enzymes to form the phytohormone abscisic acid (ABA) (Young *et al.*, 2012), which triggers closing of stomatal pores (Daszkowska-Golec and Szarejko, 2013). During the post-veraison phase, Cluster #198 (R_c = 0.46), containing osmotic-responsive transcripts, exhibited opposite trend in Sangiovese and Cabernet Sauvignon, in relation to the stomatal conductance gs and the HWI index, suggesting a different degree of resistance towards osmotic stress between the two genotypes. Notably, Clusters #279 (R_c = 0.90) and #263 (R_c = 0.82), containing members of the stilbene synthase family, scored negative correlations with the (Tx–Tm) thermal interval and heat waves index, and a positive correlation with the rainfall parameter only in the Sangiovese cultivar, suggesting this genotype has a greater capacity to produce stilbenes under favorable thermal conditions (Figure 7c, Dataset S12).

DISCUSSION

GxE studies in woody perennial plants are rare because the precise definition of the E component is often challenging in field studies (Brosché *et al.*, 2005, Holliday *et al.*, 2010, Travers *et al.*, 2010, Richards *et al.*, 2012, Dal Santo *et al.*, 2013, Dal Santo *et al.*, 2016b, Hess *et al.*, 2016). We have addressed the lack of a temporal GxE component (Grishkevich and Yanai,

2013) by providing a time-based approach for both G (fruit development) and E (vintage), given that both aspects are important in an environmentally sensitive crop such as grapevine, particularly in the context of global climate change. Our experimental design was specifically tailored to detect differences in plasticity between two grapevine genotypes (Cabernet Sauvignon and Sangiovese) cultivated in three different locations. Various parameters indicated that our sampling procedure *in field* was accurate, still our novel data-mining pipeline was designed to address the hindrance of collecting uniform developmental stages in different seasons, at different sites and in different varieties. This statistical approach comprises a three-step screening scheme to remove unwanted sources of gene expression variability, the clustering of gene co-expression profiles based on four different developmental stages, and an estimation of the inner representativeness of the clusters (i.e., the internal cohesion of each cluster). These statistical precautions allowed us to focus on the most important and consistent differences in gene expression due to the four analyzed variables, minimizing the overstating of the variability due to unforeseen differences in the collected developmental stages.

We observed a difference in transcriptomic plasticity between the two genotypes in response to the environment, which has been postulated but not empirically demonstrated in previous studies (Ortega-Regules *et al.*, 2006, Rustioni *et al.*, 2013, Zenoni *et al.*, 2017). GxE interactions became predominant during fruit maturation, particularly in Sangiovese berries. This is economically the most important phase of berry development due to the emerging aromatic profile (Conde *et al.*, 2007). The characteristics of Cabernet Sauvignon berries were less dependent on growth conditions and, accordingly, the transcriptome remained more stable across vintages and locations, suggesting that the limited plasticity may underpin the success of this cultivar in many different parts of the world. When designing the experimental layout, most of the growing conditions were set to uniformity across the three sites, but the rootstock, as Cabernet Sauvignon was grafted on three different genotypes. However, they derived from the same parent species (*V. berlandieri* x *V. riparia*), and they share similar agro/physiological characteristics (Keller, 2015). Rootstocks may have a significant impact on the interaction between plant and environment, nevertheless we observed an higher transcriptome stability in Cabernet Sauvignon across different locations than Sangiovese. This finding suggests that the rootstock did not significantly contribute to the berry transcriptome variability. This corroborates our previous findings demonstrating that environmental and growing factors have a greater impact than the rootstock on the transcriptomic plasticity in developing berries (Dal Santo *et al.*, 2013). DNA methylation analysis also revealed differences between the genotypes, suggesting that epigenetic regulation may partially explain the variation between the genotypes in terms of gene expression in different environments, as recently postulated in Shiraz cultivar (Xie *et al.*, 2017) . A recent study based on the biological material used herein has also suggested

that small RNAs have a buffering effect on transcriptomic plasticity in the widely cultivated Cabernet Sauvignon cultivar (Paim Pinto *et al.*, 2016).

We established a novel data-mining pipeline to uncover relationships among four G and E variables (stage, cultivar, vintage and location) which revealed inner hierarchies and interactions, such as vintage x location. We found that 37.5% of all modulated genes were highly canalized (i.e., expressed in a consistent profile across different genotypes and environments), representing core functions that could ultimately be developed into universal markers for berry development in the field. A further 14.6% of all modulated genes were genotype-dependent but unaffected by the environment, and were enriched in biotic stress response functions. CNV and haplotype sharing between cultivars explained some of these genotype-dependent differences in expression. The expression of a further 23.83% of the modulated genes was dependent on the vintage, location and vintage x location interaction, although the vintage and location variables *per se* showed only marginal effects on the extent of transcriptome plasticity in both genotypes (9.1% and 6.5% of the modulated genes, respectively). Indeed, this strong interaction indicated that the vintage effect (Jones and Davis, 2000, Dal Santo *et al.*, 2013, Van Leeuwen and Darriet, 2016) may have different molecular outcomes in different locations.

The pool of GxE-related genes which showed plasticity in one genotype but not the other, or different degrees of plasticity in each genotype, accounted for 9.48% of all modulated genes. Genes responsible for GxE interactions may show similar characteristics to purely genotype-dependent genes, e.g. they are often nonessential (Landry *et al.*, 2006, Tirosh *et al.*, 2006, Grishkevich and Yanai, 2013). Accordingly, we found that many grapevine GxE-related genes are involved in stress responses, signal transduction and secondary metabolism. The last of these indicates that GxE interactions may represent a point of economic leverage, particularly in specialty crops such as grapevine that are valued more for characteristics determined by secondary metabolism than for high yields. Lastly, genes related to GxE interactions showed different within-cultivar diversity in the two genotypes, supporting the hypothesis that heterozygosity may buffer against environmental variation by providing an expanded range of gene expression (Roff, 2005) and that the underlying principles governing GxE interactions are not simply the combination of factors influencing genotypic and environmental variation (Grishkevich *et al.*, 2012).

Finally, our attempt to statistically correlate gene expression data with the principal agro/physiological and meteorological/environmental parameters allowed us to retrieve those clusters of gene expression which maximized the difference between the two cultivars, in terms of the interaction with the environment. The effort to correlate large scale transcriptomic data

with such parameters, recorded in the field during the course of the experiment, could herald a modern agriculture era.

Conclusions

The new statistical pipeline described herein, combined with the observed contribution of genetic diversity to the different gene expression profiles, supports and augments previous findings (Dal Santo *et al.*, 2013). First, the transcriptomic plasticity of berries representing different locations and vintages is underpinned by broad transcriptional reprogramming. Second, within-cultivar diversity may modulate gene expression in response to environmental cues. Third, the location of the vineyard has a minor impact on the extent of GxE-dependent transcriptome plasticity in berries, but plays an important role in determining the performance of each genotype by enhancing qualitative traits such as the accumulation of secondary metabolites related to wine aroma and color.

Our study provides a multi-omics approach to separate the many layers of regulation that determine GxE interactions in field-grown plants. Given that the unprecedented rate of climate change will challenge the traditional concept of a geographically-determined *terroir* (White *et al.*, 2009), our study helps to provide a broader molecular definition of the *terroir* concept which will contribute to sustainable viticulture, wine production and marketing.

EXPERIMENTAL PROCEDURES

Description of experimental sites. Grapevine berry samples were collected from 7/10-year-old vineyards located in Bolgheri (wine cellar Podere Guado al Melo, Tuscany coast), Montalcino (wine cellar Banfi Srl, Tuscany Apennines) and Riccione (wine cellar Valbruna Soc. Coop. Agricola, Romagna coast) during the 2011–2012 growing seasons. Cabernet Sauvignon and Sangiovese berries were sampled from adjacent vineyards at each experimental site to avoid major environmental differences between cultivars (Figure 1a). The most relevant features of each vineyard are summarized in Table S1.

Meteorological data collection and analysis. The air temperature of the vineyard above the canopy layer was monitored during the 2011–2012 growing seasons at all three sites using a HOBO U23 Pro v2 thermistor thermometer (Onset Computer Corporation, Bourne, MA, USA). Daily mean (Td), daily maximum (Tx), and daily minimum (Tm) temperatures were extracted from hourly values. Daily global solar radiation (GSR) was reconstructed by applying the Hargreaves formula to Tx and Tm values (Hargreaves and Samani, 1985). The growing degree days at base 10°C (GDD_10C) was calculated by summing the average daily temperatures from June to September and subtracting 10°C per day (negative values were recorded as zero).

Rainfall data were collected from the pluviometric station nearest to each vineyard. The available water content (AWC) was estimated as previously described (Saxton and Rawls, 2006), taking into account the soil type and rainfall. For correlation analysis with transcriptomic data, the Td, Tx, Tm, Tx–Tm, GSR and rainfall parameters were also computed within the 5 days before each sampling date. The heat wave index was calculated as the sum of Tx above 30°C within two sampling dates.

Berry sampling. Berries were collected at four developmental stages: pea size (PS, 5 mm diameter, BBCH 75), pre-veraison (PV, majority of berries touching, BBCH 79), mid-ripening (MR, berries developing color, BBCH 83) and fully-ripe (FR, berries ripe for harvest, BBCH 89)(Lorenz *et al.*, 1995) at the same time of day (~11:00 am) (Figure 1b). The sampling dates are reported in Table S2. Three biological replicates of 600 berries per stage were collected from upper, central and lower parts of the cluster, and from the sun-exposed and shaded sides. The samples were divided into two groups and frozen in liquid nitrogen: 400 berries for metabolic analysis, stored at –20°C, and 200 berries for transcriptomic/epigenomic analysis, stored at –80°C.

Fruit composition and yield parameters. FR berries were harvested from six vines per variety at each site. The total soluble solids content of the pressed juice (°Brix) was determined with a refractrometer (Global Water, Sacramento, CA, USA). We also measured the pH using a pH meter (Hanna Instruments, Woonsocket, RI, USA) and titratable acidity (expressed as grams of tartaric acid per liter of juice, with 0.1 M NaOH and bromothymol blue as indicators) using an automatic titration system (Hanna Instruments). The mean berry weight was determined based on 50 berries, and we also determined the yield per vine and number of clusters per vine.

Physiological data. Gas exchange measurements were carried out at the same time of day on each sampling date (~12:00 pm). The photosynthetic rate (Pn), transpiration rate (E_T) and stomatal conductance (gs) were recorded using a CIRAS-2 portable photosynthesis system (PP Systems Ltd, Haverhill, UK). Ten stable values were recorded from different plants. The stem water potential (SWP) of non-transpiring mature leaves was monitored using a Scholander-pressure chamber (Soil Moisture Equipment Corporation, Santa Barbara, CA, USA) when the berries reached the FR stage. Ten mature, undamaged, sun-exposed leaves were selected and placed into a plastic bag wrapped with aluminum foil at least 1 h before measurement. The soil water content (SWC) at 20–40 and 60–80 cm was determined by collecting soil samples in triplicate using a soil auger, oven drying at 110°C for 24 h, and calculating the water content by comparing with the fresh weight. For correlation analysis with transcriptomic data, the mean of the 20–40 and 60–80 cm SWC values was used. The most relevant physiological data are summarized in Figure S9 (Cabernet Sauvignon) and Figure S10 (Sangiovese).

Metabolic composition of berries. The carotenoid and chlorophyll content of the berry samples was determined by high performance liquid chromatography (HPLC) as previously described (Mendes-Pinto *et al.*, 2004) with minor modifications (Kamffer *et al.*, 2010). The norisoprenoid content was determined during ripening by solid-phase micro-extraction and gas chromatography/mass spectrometry (GC-MS) as previously described (De Lorenzis *et al.*, 2017). The flavonol and HCA content was determined by HPLC as previously described (De Lorenzis *et al.*, 2017). In each case, 50 berries were used for extraction.

RNA extraction and microarray hybridization. Total RNA was extracted from approximately 400 mg berry pericarp tissue (berries without seeds) ground in liquid nitrogen, using the Spectrum™ Plant Total RNA kit (Sigma-Aldrich, St. Louis, MO, USA)(Dal Santo *et al.*, 2016a). We hybridized 5 µg of total RNA per sample to a NimbleGen microarray 090818_Vitus_exp_HX12 chip (Roche, NimbleGen Inc., Madison, WI, USA) containing probes representing 29,549 predicted grapevine genes covering ~98.6% of the genes predicted in the V1 annotation of the 12X grapevine genome. Each microarray was scanned using an Axon GenePix 4400A (Molecular Devices, Sunnyvale, CA, USA) at 532 nm (Cy3 absorption peak) and GenePix Pro7 software (Molecular Devices) according to the manufacturer's instructions. Images were analyzed using NimbleScan v2.5 software (Roche), which produces Pair Files containing the raw signal intensity data for each probe and Calls Files with normalized expression data derived from the average of the intensities of the four probes for each gene.

Statistical analysis of microarray data. Correlation matrixes were prepared using R software and Pearson's correlation coefficient as the statistical metric to compare the values of the whole transcriptome in all samples using the average value of the three biological replicates. Correlation values were converted into distance coefficients to define the height scale of the dendrogram. A non-parametric Kruskall-Wallis test (FDR=0.01%, 24 classes, Benjamini and Hochberg correction) was applied to each of two 72-sample genotype-specific datasets. After assessing the unimodal distribution of the fluorescent intensities (Fasoli *et al.*, 2012, Dal Santo *et al.*, 2013) (Figure S2) with R software, a two-sided two-way ANOVA (1,000 permutations, $p<0.01$, vintage and location classes) was applied to each dataset using TMeV v4.8.

Correlation between transcriptomic and climatic/agricultural data. Correlation matrixes were prepared using Spearman's correlation coefficient in R software to compare trends in the mean expression values of each of the 300 clusters (Dataset S3) with the trends of climatic and agricultural parameters. A first general matrix compared 48 conditions (two cultivars, four stages, three locations and two vintages) whereas four genotype-specific 12-condition matrices were prepared for the separate analysis of pre-veraison and post-veraison samples (one cultivar, two stages, three locations and two vintages). Subtraction matrices were generated for

the latter Sangiovese and Cabernet Sauvignon correlation matrices. The mathematical operation was performed only on Spearman's correlation values ≥0.6, and only subtraction values ≥|0.65| were considered biologically relevant.

Design of a statistical pipeline to inspect GxE interactions using microarray data. A detailed description of the statistical pipeline is provided in Text S2. A Venn diagram was prepared using the top 100 scoring clusters in each variable's VIM ranking (Dataset S2) using Venny v2.1 (http://bioinfogp.cnb.csic.es/tools/venny/). Gene Ontology (GO) annotation was applied using the BiNGO v2.3 plug-in tool in Cytoscape v2.6 with PlantGOslim categories (Maere *et al.*, 2005). Overrepresented PlantGOslim categories were identified using a hypergeometric test with a significance threshold of 0.05. Bar plots ranking, when possible, the top five biological processes were prepared based on enrichment scores [$-\log_{10}$ (P value)].

RNA-seq and data analysis. The PV and MR triplicate samples (two cultivars, three locations and two vintages) yielded 72 non-directional cDNA libraries, which were prepared from 2.5 µg of total RNA using the Illumina TruSeq RNA Sample preparation protocol (Illumina Inc., San Diego CA, USA) according to the manufacturer's instructions. Single-end reads of 100 nucleotides were obtained using an Illumina Hiseq 2000 sequencer, and sequencing data were generated using the base-calling software Illumina Casava v1.8.2 (31,091,566 ± 6,162,118 reads per sample). The reads were aligned onto the PN40024 12X reference genome (Jaillon *et al.*, 2007) using TopHat v2.0.9 (Kim *et al.*, 2013) with default parameters. An average of 86.91% of reads was mapped for each sample (Table S8). Transcripts were assembled from mapped reads and normalized transcript abundance measurements expressed in FPKM values were prepared using Cufflinks v2.1.1 (Trapnell *et al.*, 2010) resulting in a non-redundant list of 29,971 transcripts.

Reduced representation bisulfite sequencing (RRBS) and data analysis. The PV and MR duplicate samples (two cultivars, three locations and two vintages) were used to prepare 48 RRBS libraries as previously reported, with modifications (Gu *et al.*, 2011). Briefly, 200 ng of genomic DNA was digested with *Taq*I (NEB, Ipswich, MA, USA) at 65°C for 2 h. After purification using the QIAquick PCR Purification kit (Qiagen, Hilden, Germany), fragment ends were repaired and ligated using adapters provided in the Ovation Ultralow Methyl-Seq DR Multiplex Kit (NuGEN, San Carlos, CA, USA). Ligated products corresponding to 100–1500 bp DNA fragments were purified by 2% low-range agarose gel electrophoresis before final end-repair using the same NuGEN kit. Bisulfite conversion was conducted using the EpiTect Fast DNA Bisulfite Kit (Qiagen). The final RRBS libraries were generated by PCR and validated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Libraries were sequenced using the Illumina HiSeq2500 platform in paired-end 125-bp runs. Raw sequencing

data quality was evaluated using FastQC software (Babraham Institute, Cambridge, UK). Adaptor sequences were removed using TRIM GALORE (Babraham Institute) with default settings and hard-trimmed from position 1–5 nt to improve data quality. Cleaned reads were aligned to the grapevine reference genome (Jaillon *et al.*, 2007) using the bisulfite alignment program Bismark v0.14.5 (Krueger and Andrews, 2011) yielding an average of ~15 million read pairs uniquely aligned per sample. Alignments were deduplicated and converted into single-cytosine methylation maps using the Bismark package with default settings. In total, ~975,000 CG sites, ~1 million CHG sites and ~5.8 million CHH sites were covered by at least one read on average per sample. Cytosine positions identified as C→T or G→A polymorphisms were discarded to remove false bisulfite conversion signals and remaining cytosine residues were filtered by minimum coverage in all 48 samples with different thresholds depending on sequence context (CG = 4, CHG =10 and CHH = 10). The final set of cytosine residues was analyzed separately by context using the methylKit R package (Akalin *et al.*, 2012), which identified 4,696 CG sites, 4,737 CHG sites and 14,179 CHH sites that could be compared among all the 48 samples. Analysis of differential methylation was based on logistic regression, and k-means and unscaled PCA were applied to the set of shared CG, CHG and CHH sites using the R functions kmeans() and prcomp(), respectively. Significant associations between principal components and experimental covariates (biological replicate, vintage, cultivar, developmental stage and location) were identified using a Pearson's correlation test.

Haplotype sharing. Genomic DNA from each cultivar was sequenced on Illumina HiSeq2500 sequencing apparatus to produce 2 x 100 paired end reads that were aligned to the 12X V0 version of the grapevine reference genome (Jaillon *et al.*, 2007) using BWA (Li and Durbin, 2009) with default parameters. Single nucleotide polymorphisms (SNPs) were called using GATK Unified Genotyper variant discovery (McKenna *et al.*, 2010, DePristo *et al.*, 2011). SNPs with phred-scaled quality score < 50 or minimum coverage < 5 reads or read coverage ≤ 0.5 or ≥ 1.5 X of the modal coverage were removed. Heterozygous genotypes were called when the reference/alternate allele ratio was ≥ 0.25 and ≤ 0.75. Haplotype sharing was computed in 2,367 genome windows of 100 kb of putatively single copy DNA, obtained after masking transposable elements and other repeats. IBD in each genome window was calculated using a slightly modified version of the identity-by-state ratio method used in citrus (Wu *et al.*, 2014) with the following thresholds: IBD=0 if IBSR < 0.95 and D > 0.025; IBD=1 if IBSR ≥ 0.95 and D > 0.025; IBD=2 if IBSR ≥ 0.95 and D ≤ 0.025. IBSR and D were calculated with following formulas: IBSR = (IBS2+IBS1)/(IBS2+IBS1+IBS0); D=[(IBS1*0.5)+IBS0]/(IBS0+IBS1+IBS2+no. of invariant sites).. We defined subsets of homozygous or heterozygous genes based on SNP frequencies in the predicted transcribed

pèortion of the gene, and up to 2 kb upstream of the start site. We estimated an error rate of 0.004 heterozygous SNP calls in genes located in genomic windows with complete haplotype sharing between PN40024 and Cabernet Sauvignon/Sangiovese. We therefore classified as homozygous all genes with <0.004 heterozygous SNPs per mappable site. The remaining genes were classified as heterozygous. Windows containing centromeric repeats and adjacent windows with >50% repetitive DNA were classified as pericentromeric regions. All other windows were assigned to chromosome arms.

Copy number variants. Depth of coverage was analyzed in non-overlapping windows of variable size, containing a constant number of 1,500 mappable reads. To define these windows, wgsim (https://github.com/lh3/wgsim) was used to simulate 100 million 100-bp long reads from the grapevine reference genome, with a mean insert size of 500 bp (Jaillon *et al.*, 2007). Simulated reads were aligned to the reference genome using BWA (Li and Durbin, 2009) with default parameters and duplicated sequences were removed with the samtools rmdup utility (Li *et al.*, 2009). The number of uniquely mapped paired reads was used to define window sizes. The average window size for 1,500 mappable reads was 4.6 kb. In each window, we calculated the log2 ratio between the number of mapped reads in the reference genome and the number of mapped reads in the Cabernet Sauvignon or Sangiovese genomes. The ratios were normalized on the basis of the total number of paired reads mapped in each variety and were used as an input for the binary circular segmentation implemented in DNAcopy (Olshen *et al.*, 2004). The R package edgeR (Robinson *et al.*, 2010) was used to estimate the significance of the log2 ratio in each window within the segments identified by DNAcopy. Segments with a median significance <0.05 were selected as CNVs. Segments with a log2 ratio of 0.5–2.5 were classified as hemizygous, and those with a log2 ratio >2.5 were classified as deleted. Across the 19 grapevine chromosomes, 39.45 and 35.41 Mb of genomic DNA were affected by CNVs in Cabernet Sauvignon and Sangiovese, respectively.

Statistical analysis. The yield, fruit composition, HPLC and GC-MS data were analyzed using SPSS statistical software vPASW Statistics 22 (SPSS, Inc., Chicago, IL, USA). ANOVA was used to test the main effects (cultivar, location and vintage) and their interactions. Means were compared using Duncan's test at $p < 0.05$. The data were plotted using SigmaPlot software v11 (Systat Software, Inc., San Jose, CA). A chi-squared test was used to compare genomic distribution frequencies (χ^2>0.01 unless otherwise specified).

Accession numbers and data availability

The datasets supporting the conclusions of this article are available in the following repositories. All microarray expression data are available at GEO under the series entry GSE97578 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=idanyawwdjeppwn&acc=GSE97578). All RNA-seq data are available at GEO under the series entry GSE97960 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=cpgfuqewbjsjnaz&acc=GSE97960). The RRBS data are available at GEO under the series entry GSE98762 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=otutoigcddgzbkb&acc=GSE98762). The genome sequences of Sangiovese and Cabernet Sauvignon are available at NCBI, BioProject ID SRP106422.

ACKNOWLEDGEMENTS

We thank Dr. Paolo Sabbatini for critically reading the manuscript. We also thank Ms. Nicoletta Felice for RRBS library construction. This work was supported by the 'Valorizzazione dei Principali Vitigni Autoctoni Italiani e dei loro Terroir (Vigneto)' project funded by the Italian Ministry of Agricultural and Forestry Policies and partly by the European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013) [Grant Agreement number 294780, Novabreed]. SDS and CDF were financed by the Italian Ministry of University and Research FIRB RBFR13GHC5 project "The Epigenomic Plasticity of Grapevine in Genotype per Environment Interactions". The authors declare no conflict of interest.

AUTHORS CONTRIBUTIONS

S.D.S. performed the transcriptome experiments, analyzed the data, interpreted the results, coordinated the scientific project and wrote the manuscript; S.Z. analyzed the data, interpreted the results and wrote the manuscript; M.S. designed the statistical pipeline, performed the data mining procedures on data and drafted the manuscript; G.D.L., L.B. and M.R. sampled the biological material, performed the eco-physiological characterization and drafted the manuscript; G.M., G.D.G. and M.M. sequenced the two genotypes, performed genomic studies and drafted the manuscript, E.D.P and C.D.F. developed DNA libraries, performed RRBS analysis and drafted the manuscript; M.F. helped with microarray analysis; P.Z. designed the statistical pipeline, supervised the data mining procedures and reviewed the manuscript; G.B.T. designed the experimental plan, helped in interpreting the results and in writing the manuscript, M.P.

conceived and supervised the study, reviewed the manuscript. All authors read and approved the final manuscript.

SHORT SUPPORTING INFORMATION LEGENDS

Figure S1. Eco-physiological characterization of the vineyards included in this study.

Figure S2. Distribution of fluorescence intensity in the genotype-specific 72-sample dataset.

Figure S3. Characterization of reduced representation bisulfite sequencing (RRBS) data.

Figure S4. Examples of genes with uninteresting expression patterns screened out from statistical analysis according to the guidelines in Table S7.

Figure S5. Variance accounted for (VAF) by clustering vs. number of clusters and examples of clusters with high and low homogeneity indices (R_c).

Figure S6. Selected two-dimensional object scores plots of the 48 experimental conditions.

Figure S7. Characterization of the variable-shared clusters of gene expression.

Figure S8. Differences in transcriptomic plasticity between the two genotypes in the interaction with the environment.

Figure S9. Physiological parameters of Cabernet Sauvignon during the 2011 and 2012 growing seasons collected at four different developmental stages in the three locations.

Figure S10. Physiological parameters of Sangiovese during the 2011 and 2012 growing seasons collected at four different developmental stages in the three locations.

Table S1. Main agronomic features of the Cabernet Sauvignon and Sangiovese vineyards in Bolgheri, Montalcino and Riccione (Italy).

Table S2. Sampling dates of Cabernet Sauvignon and Sangiovese berries during the 2011 and 2012 seasons in three locations at different berry developmental stages.

Table S3. Description of sample names used for the present study.

Table S4. Main agronomical and ripening parameters.

Table S5. Physiological and biochemical characterization of Cabernet Sauvignon grapes.

Table S6. Physiological and biochemical characterization of Sangiovese grapes.

Table S7. Screening guidelines used to remove genes with uninteresting expression profiles.

Table S8. Summary of RNA-seq data and mapping metrics.

Text S1. Environmental and varietal characterization of the samples used for the present work.

Text S2. Detailed description of the new statistical pipeline.

Text S3. Characterization of the variable-specific and variable –shared clusters of gene expression.

Dataset S1. Set of the 18,122 modulated genes analyzed using the statistical pipeline.

Dataset S2. Characterization of the 300 clusters of GxE gene expression.

Dataset S3. Detailed description of all the 300 clusters of gene expression defined in the present work.

Dataset S4. Characterization of the variable-specific clusters.

Dataset S5. List of genes sorted by their intrachromosomal location.

Dataset S6. Copy number variation (CNV) in genes belonging to the 300 clusters.

Dataset S7. Characterization of the variable-shared clusters.

Dataset S8. Haplotype sharing between the Cabernet Sauvignon and Sangiovese genotypes.

Dataset S9. List of genes sorted by the zygosity of the corresponding locus in Cabernet Sauvignon and Sangiovese.

Dataset S10. Spearman's correlation matrix between transcriptomic and climatic/physiological data over the whole dataset (48 samples).

Dataset S11. Spearman's correlation matrices for transcriptomic and climatic/physiological data in pre-veraison (PS and PV) and post-veraison (MR and FR) samples for the Cabernet Sauvignon and Sangiovese genotypes, separately.

Dataset S12. Subtraction matrices of the Spearman's correlation matrices obtained for the pre-veraison and post-veraison samples in the two genotypes.

REFERENCES

Adam-Blondon, A.F., Martinez-Zapater, J.M. and Kole, C. (2011) *Genetics, Genomics, and Breeding of Grapes*: CRC Press.

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A. and Mason, C.E. (2012) methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome biology*, **13**.

Anesi, A., Stocchero, M., Dal Santo, S., Commisso, M., Zenoni, S., Ceoldo, S., Tornielli, G.B., Siebert, T.E., Herderich, M., Pezzotti, M. and Guzzo, F. (2015) Towards a scientific

interpretation of the terroir concept: plasticity of the grape berry metabolome. *BMC plant biology*, **15**, 191.

Brosché, M., Vinocur, B., Alatalo, E.R., Lamminmäki, A., Teichmann, T., Ottow, E.A., Djilianov, D., Afif, D., Bogeat-Triboulot, M.-B., Altman, A., Polle, A., Dreyer, E., Rudd, S., Paulin, L., Auvinen, P. and Kangasjärvi, J. (2005) Gene expression and metabolite profiling of Populus euphratica growing in the Negev desert. *Genome biology*, **6**, R101.

Conde, C., Silva, P., Fontes, N., Dias, A.C.P., Tavares, R.M., Sousa, M.J., Agasse, A., Delrot, S. and Gerós, H. (2007) Biochemical changes throughout grape berry development and fruit and wine quality. *Food*.

Dai, Z.W., Ollat, N., Gomès, E., Decroocq, S., Tandonnet, J.-P., Bordenave, L., Pieri, P., Hilbert, G., Kappel, C., van Leeuwen, C., Vivin, P. and Delrot, S. (2011) Ecophysiological, Genetic, and Molecular Causes of Variation in Grape Berry Weight and Composition: A Review. *American Journal of Enology and Viticulture*, **62**, 413-425.

Dal Santo, S., Commisso, M., D'Inc, Erica, Anesi, A., Stocchero, M., Zenoni, S., Ceoldo, S., Tornielli, G.B., Pezzotti, M. and Guzzo, F. (2016a) The Terroir Concept Interpreted through Grape Berry Metabolomics and Transcriptomics. *JoVE*, e54410.

Dal Santo, S., Fasoli, M., Negri, S., D'Inca, E., Vicenzi, N., Guzzo, F., Tornielli, G.B., Pezzotti, M. and Zenoni, S. (2016b) Plasticity of the Berry Ripening Program in a White Grape Variety. *Frontiers in plant science*, **7**, 970.

Dal Santo, S., Tornielli, G.B., Zenoni, S., Fasoli, M., Farina, L., Anesi, A., Guzzo, F., Delledonne, M. and Pezzotti, M. (2013) The plasticity of the grapevine berry transcriptome. *Genome biology*, **14**, r54.

Daszkowska-Golec, A. and Szarejko, I. (2013) Open or close the gate - stomata action under the control of phytohormones in drought stress conditions. *Frontiers in plant science*, **4**.

De Lorenzis, G., Squadrito, M., Rossoni, M., di Lorenzo, G.S., Brancadoro, L. and Scienza, A. (2017) Study of intra-varietal diversity in biotypes of Aglianico and Muscat of Alexandria (Vitis vinifera L.) cultivars. *Australian Journal of Grape and Wine Research*, **23**, 132-142.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D. and Daly, M.J. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*, **43**, 491-+.

Des Marais, D.L., Hernandez, K.M. and Juenger, T.E. (2013) Genotype-by-Environment Interaction and Plasticity: Exploring Genomic Responses of Plants to the Abiotic Environment. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 5-29.

El-Soda, M., Malosetti, M., Zwaan, B.J., Koornneef, M. and Aarts, M.G.M. (2014) Genotype x environment interaction QTL mapping in plants: lessons from Arabidopsis. *Trends Plant Sci*, **19**, 390-398.

Fasoli, M., Dal Santo, S., Zenoni, S., Tornielli, G.B., Farina, L., Zamboni, A., Porceddu, A., Venturini, L., Bicego, M., Murino, V., Ferrarini, A., Delledonne, M. and Pezzotti, M.

(2012) The Grapevine Expression Atlas Reveals a Deep Transcriptome Shift Driving the Entire Plant into a Maturation Program. *Plant Cell*, **24**, 3489-3505.

Fournier-Level, A., Korte, A., Cooper, M.D., Nordborg, M., Schmitt, J. and Wilczek, A.M. (2011) A Map of Local Adaptation in *Arabidopsis thaliana*. *Science*, **334**, 86-89.

Friedman, J.H. (2001) Greedy function approximation: A gradient boosting machine. *Ann Stat*, **29**, 1189-1232.

Grishkevich, V., Ben-Elazar, S., Hashimshony, T., Schott, D.H., Hunter, C.P. and Yanai, I. (2012) A genomic bias for genotype-environment interactions in C. elegans. *Mol Syst Biol*, **8**.

Grishkevich, V. and Yanai, I. (2013) The genomic determinants of genotype x environment interactions in gene expression. *Trends Genet*, **29**, 479-487.

Gu, H.C., Smith, Z.D., Bock, C., Boyle, P., Gnirke, A. and Meissner, A. (2011) Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc*, **6**, 468-481.

Hargreaves, G.H. and Samani, Z.A. (1985) Reference crop evapotranspiration from temperature. *Transactions of the ASAE*, **1**, 96-99.

Hess, M., Wildhagen, H., Junker, L.V. and Ensminger, I. (2016) Transcriptome responses to temperature, water availability and photoperiod are conserved among mature trees of two divergent Douglas-fir provenances from a coastal and an interior habitat. *BMC genomics*, **17**, 682.

Holliday, J.A., Ritland, K. and Aitken, S.N. (2010) Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (Picea sitchensis). *New Phytologist*, **188**, 501-514.

Jaillon, O., Aury, J.M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyere, C., Billault, A., Segurens, B., Gouyvenoux, M., Ugarte, E., Cattonaro, F., Anthouard, V., Vico, V., Del Fabbro, C., Alaux, M., Di Gaspero, G., Dumas, V., Felice, N., Paillard, S., Juman, I., Moroldo, M., Scalabrin, S., Canaguier, A., Le Clainche, I., Malacrida, G., Durand, E., Pesole, G., Laucou, V., Chatelet, P., Merdinoglu, D., Delledonne, M., Pezzotti, M., Lecharny, A., Scarpelli, C., Artiguenave, F., Pe, M.E., Valle, G., Morgante, M., Caboche, M., Adam-Blondon, A.F., Weissenbach, J., Quetier, F., Wincker, P. and Public, F.-I. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, **449**, 463-467.

Jones, G.V. and Davis, R.E. (2000) Climate Influences on Grapevine Phenology, Grape Composition, and Wine Production and Quality for Bordeaux, France. *American Journal of Enology and Viticulture*, **51**, 249-261.

Kamffer, Z., Bindon, K.A. and Oberholster, A. (2010) Optimization of a Method for the Extraction and Quantification of Carotenoids and Chlorophylls during Ripening in Grape Berries (Vitis vinifera cv. Merlot). *J Agr Food Chem*, **58**, 6578-6586.

Keller, M. (2010) Managing grapevines to optimise fruit development in a challenging environment: a climate change primer for viticulturists. *Australian Journal of Grape and Wine Research*, **16**, 56-69.

Keller, M. (2015) *The Science of Grapevines* Second edn. Oxford, UK: Academic Press, Elsevier Inc.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*, **14**.

Krueger, F. and Andrews, S.R. (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, **27**, 1571-1572.

Landry, C.R., Oh, J., Hartl, D.L. and Cavalieri, D. (2006) Genome-wide scan reveals that genetic variation for transcriptional plasticity in yeast is biased towards multi-copy and dispensable genes. *Gene*, **366**, 343-351.

Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Proc, G.P.D. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.

Lorenz, D.H., Eichhorn, K.W., Bleiholder, H., Klose, R., Meier, U. and Weber, E. (1995) Growth Stages of the Grapevine: Phenological growth stages of the grapevine (*Vitis vinifera* L. ssp. *vinifera*). *Australian Journal of Grape and Wine Research*, **1**, 100-103.

Maere, S., Heymans, K. and Kuiper, M. (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, **21**, 3448-3449.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. and DePristo, M.A. (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297-1303.

Mendes-Pinto, M.M., Ferreira, A.C.S., Oliveira, M.B.P.P. and DE Pinho, P.G. (2004) Evaluation of some carotenoids in grapes by reversed- and normal-phase liquid chromatography: A qualitative analysis. *J Agr Food Chem*, **52**, 3182-3188.

Nicotra, A.B., Atkin, O.K., Bonser, S.P., Davidson, A.M., Finnegan, E.J., Mathesius, U., Poot, P., Purugganan, M.D., Richards, C.L., Valladares, F. and van Kleunen, M. (2010) Plant phenotypic plasticity in a changing climate. *Trends Plant Sci*, **15**, 684-692.

Olshen, A.B., Venkatraman, E.S., Lucito, R. and Wigler, M. (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*, **5**, 557-572.

Ortega-Regules, A., Romero-Cascales, I., Lopez-Roca, J.M., Ros-Garcia, J.M. and Gomez-Plaza, E. (2006) Anthocyanin fingerprint of grapes: environmental and genetic variations. *J Sci Food Agr*, **86**, 1460-1467.

Paim Pinto, D.L., Brancadoro, L., Dal Santo, S., De Lorenzis, G., Pezzotti, M., Meyers, B.C., Pe, M.E. and Mica, E. (2016) The Influence of Genotype and Environment on Small RNA Profiles in Grapevine Berry. *Frontiers in plant science*, **7**, 1459.

Perry, J.C. and Mank, J.E. (2014) From Genotype x Environment to Transcriptome x Environment: Identifying and Understanding Environmental Influences in the Gene Expression Underlying Sexually Selected Traits. In *Genotype-by-Environment Interactions and Sexual Selection* (Hunt, J. and Hosken, D. eds). Hoboken, NJ: John Wiley & Sons, Ltd, pp. 169-188.

Pourcel, L., Routaboul, J.M., Kerhoas, L., Caboche, M., Lepiniec, L. and Debeaujon, I. (2005) TRANSPARENT TESTA10 encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in Arabidopsis seed coat. *Plant Cell*, **17**, 2966-2980.

Richards, C.L., Rosas, U., Banta, J., Bhambhra, N. and Purugganan, M.D. (2012) Genome-wide patterns of Arabidopsis gene expression in nature. *PLoS genetics*, **8**, e1002662.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139-140.

Rockman, M.V. and Kruglyak, L. (2006) Genetics of global gene expression. *Nat Rev Genet*, **7**, 862-872.

Roff, D.A. (2005) Variation and Life-History Evolution. In *Variation: A Central Concept in Evolution* (Hallgrímsson, B. and Hall, B.K. eds). Burlington, MA, USA: Elsevier Academic Press, pp. 333-357.

Rustioni, L., Rossoni, M., Failla, O. and Scienza, A. (2013) Anthocyanin Esterification in Sangiovese Grapes. *Ital J Food Sci*, **25**, 133-141.

Sasaki, E., Zhang, P., Atwell, S., Meng, D. and Nordborg, M. (2015) "Missing" G x E Variation Controls Flowering Time in Arabidopsis thaliana. *PLoS genetics*, **11**, e1005597.

Saxton, K.E. and Rawls, W.J. (2006) Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. *Soil Sci. Soc. Am. J.*, **70**, 1569-1578.

Thomas, D. (2010) Methods for Investigating Gene-Environment Interactions in Candidate Pathway and Genome-Wide Association Studies. *Annual Review of Public Health*, **31**, 21-36.

Tirosh, I., Weinberger, A., Carmi, M. and Barkai, N. (2006) A genetic signature of interspecies variations in gene expression. *Nature genetics*, **38**, 830-834.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J. and Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*, **28**, 511-515.

Travers, S.E., Tang, Z., Caragea, D., Garrett, K.A., Hulbert, S.H., Leach, J.E., Bai, J., Saleh, A., Knapp, A.K., Fay, P.A., Nippert, J., Schnable, P.S. and Smith, M.D. (2010) Variation in gene expression of Andropogon gerardii in response to altered environmental conditions associated with climate change. *Journal of Ecology*, **98**, 374-383.

Van Leeuwen, C. and Darriet, P. (2016) The Impact of Climate Change on Viticulture and Wine Quality. *Journal of Wine Economics*, **11**, 150-167.

White, M.A., Whalen, P. and Jones, G.V. (2009) Land and wine. *Nat Geosci*, **2**, 82-84.

Wu, G.A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., Perrier, X., Ruiz, M., Scalabrin, S., Terol, J., Takita, M.A., Labadie, K., Poulain, J., Couloux, A., Jabbari, K., Cattonaro, F., Del Fabbro, C., Pinosio, S., Zuccolo, A., Chapman, J., Grimwood, J., Tadeo, F.R., Estornell, L.H., Munoz-Sanz, J.V., Ibanez, V., Herrero-Ortega, A., Aleza, P., Perez-Perez, J., Ramon, D., Brunel, D., Luro, F., Chen, C.X., Farmerie, W.G., Desany, B., Kodira, C., Mohiuddin, M., Harkins, T., Fredrikson, K., Burns, P., Lomsadze, A., Borodovsky, M., Reforgiato, G., Freitas-Astua, J., Quetier, F., Navarro, L., Roose, M., Wincker, P., Schmutz, J., Morgante, M., Machado, M.A., Talon, M., Jaillon, O., Ollitrault, P., Gmitter, F. and Rokhsar, D. (2014) Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nat Biotechnol*, **32**, 656-+.

Xie, H., Konate, M., Sai, N., Tesfamicael, K.G., Cavagnaro, T., Gilliham, M., Breen, J., Metcalfe, A., Stephen, J.R., De Bei, R., Collins, C. and Lopez, C.M.R. (2017) Global DNA Methylation Patterns Can Play a Role in Defining Terroir in Grapevine (Vitis vinifera cv. Shiraz). *Frontiers in plant science*, **8**.

Yang, T.B. and Poovaiah, B.W. (2003) Calcium/calmodulin-mediated signal network in plants. *Trends Plant Sci*, **8**, 505-512.

Young, P.R., Lashbrooke, J.G., Alexandersson, E., Jacobson, D., Moser, C., Velasco, R. and Vivier, M.A. (2012) The genes and enzymes of the carotenoid metabolic pathway in Vitis vinifera L. *BMC genomics*, **13**.

Zenoni, S., Dal Santo, S., Tornielli, G.B., D'Incà, E., Filippetti, I., Pastore, C., Allegro, G., Silvestroni, O., Lanari, V., Pisciotta, A., Di Lorenzo, R., PALLIOTTI, A., Tombesi, S., Gatti, M. and Poni, S. (2017) Transcriptional responses to pre-flowering leaf defoliation in grapevine berry from different growing sites, years and genotypes. *Frontiers in plant science*, **8**.

FIGURE LEGENDS

Figure 1. Eco-physiological characterization. **(A)** Geographic locations and climatic trends of the vineyards investigated in this study. CS, Cabernet Sauvignon; SG, Sangiovese. Td(°C), daily mean temperature; GSR, daily global solar radiation; AWC, available water content. Yellow line, data collected during the 2011 season; Purple line, data collected during the 2012 season. **(B)** Physiological characterization of the sampled berries. Upper panel, the four berry developmental stages analyzed in the study over the double-sigmoid grapevine berry ripening curve. PS, pea size; PV, pre-veraison; MR, mid-ripening and FR, fully ripe. Lower panel, mean berry weight at each time point, for Cabernet Sauvignon (red) and Sangiovese (blue) in the three locations (different textures). The mean values of total soluble solids (°Brix) refer to the FR stage (maximum sugar accumulation). Bars show mean values ± SD (n=50); different letters

indicate significant differences among sites according to Duncan's test at $p<0.05$. See Text S1 for more details of the eco-physiological characterization.

Figure 2. Unsupervised analysis of the transcriptomic plasticity and methylation status of Cabernet Sauvignon and Sangiovese berries. Pearson's distance correlation matrix **(A)** and cluster dendrogram **(B)** to compare the transcriptomes of each sample, based on the average expression value of the three biological replicates. The left side bar indicates the consistency of the berry transcriptome among three locations for the two genotypes (red = Cabernet Sauvignon, blue = Sangiovese; changing bar texture represents inconsistency among transcriptomes in the three locations). Sample names are based on genotype (CS, Cabernet Sauvignon; SG, Sangiovese) followed by location (MO, Montalcino; BO, Bolgheri; RI, Riccione), developmental stage (PS, pea size, dark green; PV, pre-veraison, light green; MR, mid-ripening, slate blue, and FR, fully ripe, dark blue) and vintage (11 = 2011, 12 = 2012). See Table S3 for more details. **(C)** Transcriptomic plasticity differs in the two genotypes. Analysis of variance (two-sided two-way ANOVA, $p<0.01$, vintage and location classes) was computed on each of the two genotype-specific datasets. The number of differentially expressed genes per variable is shown. **(D-F)** Differentially methylated regions (DMRs) define the two genotypes. PCA scatter plots of DMR values were obtained separately for the three methylation contexts: CG and CHG, first two components; CHH, first and fifth components. Red = Cabernet Sauvignon, blue = Sangiovese.

Figure 3. A novel statistical pipeline defines hierarchies among experimental variables. **(A)** Schematic diagram illustrating the three-step statistical pipeline. See Text S2 for more details. **(B-E)** Description of the genotypic (stage and cultivar) and environmental (vintage and location) variable-related cluster of expression. Scatterplot of the 300 clusters according to the rank in **(B)** VIM^{c}_{Stage}, **(C)** $VIM^{c}_{Cultivar}$, **(D)** $VIM^{c}_{Vintage}$, and **(E)** $VIM^{c}_{Location}$ (i.e., RnkVIMStage = rank of clusters according to VIM^{c}_{Stage}; low values denote high importance of stage) and to the loading in the specific rotated principal component (DimRot) (first, second, third and tenth component for stage, cultivar, vintage and location, respectively). Each dot represents a single cluster, colored according to the cluster homogeneity index, R_c. Relevant examples of variable-specific clusters are given at the side of each scatter plot. See Dataset S3 for a complete description of the 300 clusters.

Figure 4. Characterization of variable-specific clusters. **(A)** Venn diagram showing the number of variable-specific and variable-shared clusters. **(B)** Summary of the principal properties of each group of clusters. Dataset S2, Dataset S3 and Dataset S4 provide a complete description of each cluster. **(C-E)** Characterization of the stage-specific clusters. **(C)** Bar plot ranking of the top five biological processes based on Gene Ontology (GO) enrichment scores within the stage-

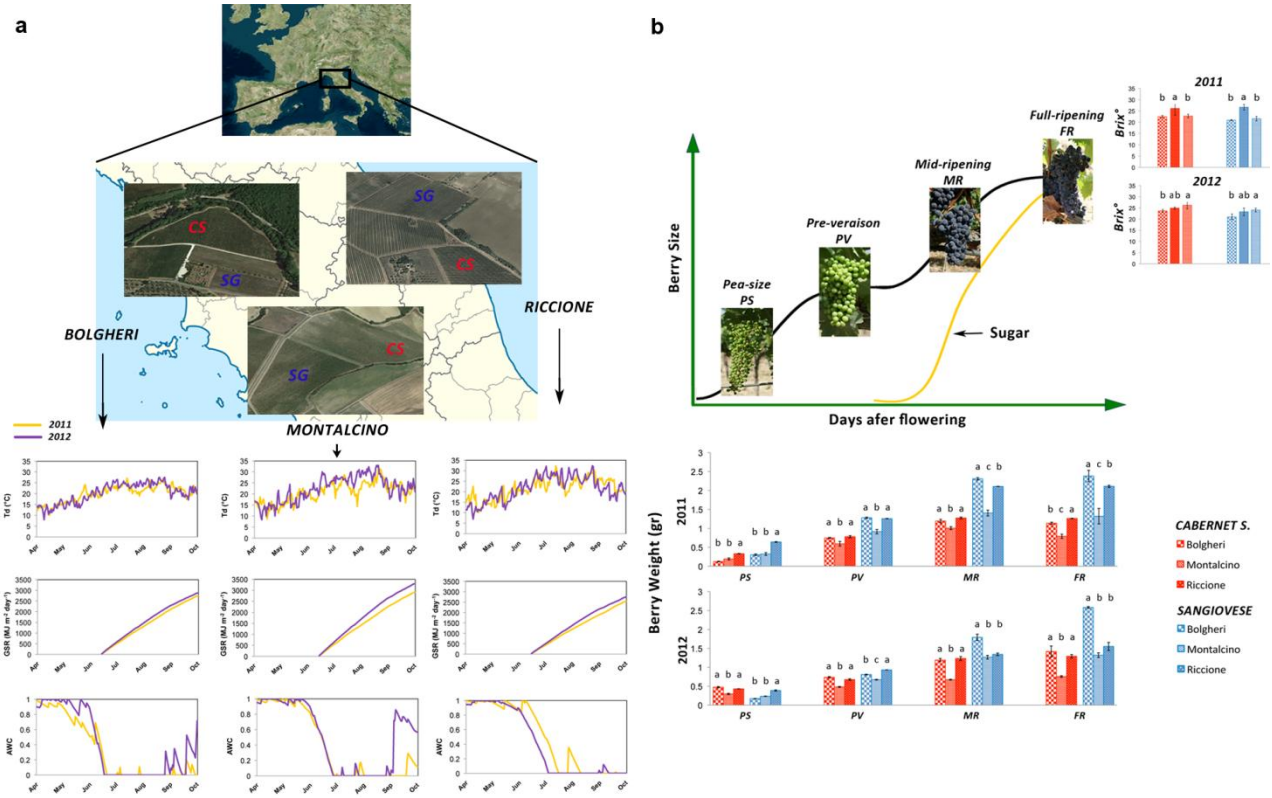
specific cluster genes. **(D)** Analysis of the expression patterns of the stage-specific cluster genes. The concavity Stage L parameter (Dataset S2) indicates the upregulation (red) or downregulation (green) expression trend. **(E)** Genome-wide distribution of stage-specific genes (white-blue) and all genes (white-black) in 100-kb windows of non-repetitive DNA. Black dots indicate the site of centromeric repeat sequence. **(F-G)** Characterization of the cultivar-specific clusters. **(F)** Bar plot ranking of the top five biological processes based on GO enrichment score within the cultivar-specific cluster genes. **(G)** Box plot of transcript levels of genes with copy number variations (left graph, genes absent from Sangiovese; right graph, genes absent from Cabernet Sauvignon). Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by *R* software; whiskers extend to 5th and 95th percentiles, outliers are represented by dots; crosses represent sample means (n = 12 left panel, n = 19 right panel). **(H)** Bar plot ranking of the top five biological processes based on GO enrichment score within the vintage-specific cluster genes. **(I)** Bar plot ranking of the two enriched biological processes based on GO enrichment scores within the location-specific cluster genes. The enriched GO biological processes were identified and listed according to their enrichment p-value ($p<0.05$). The total number of GO category-related genes within the analyzed genes query is shown on the side of each bar.

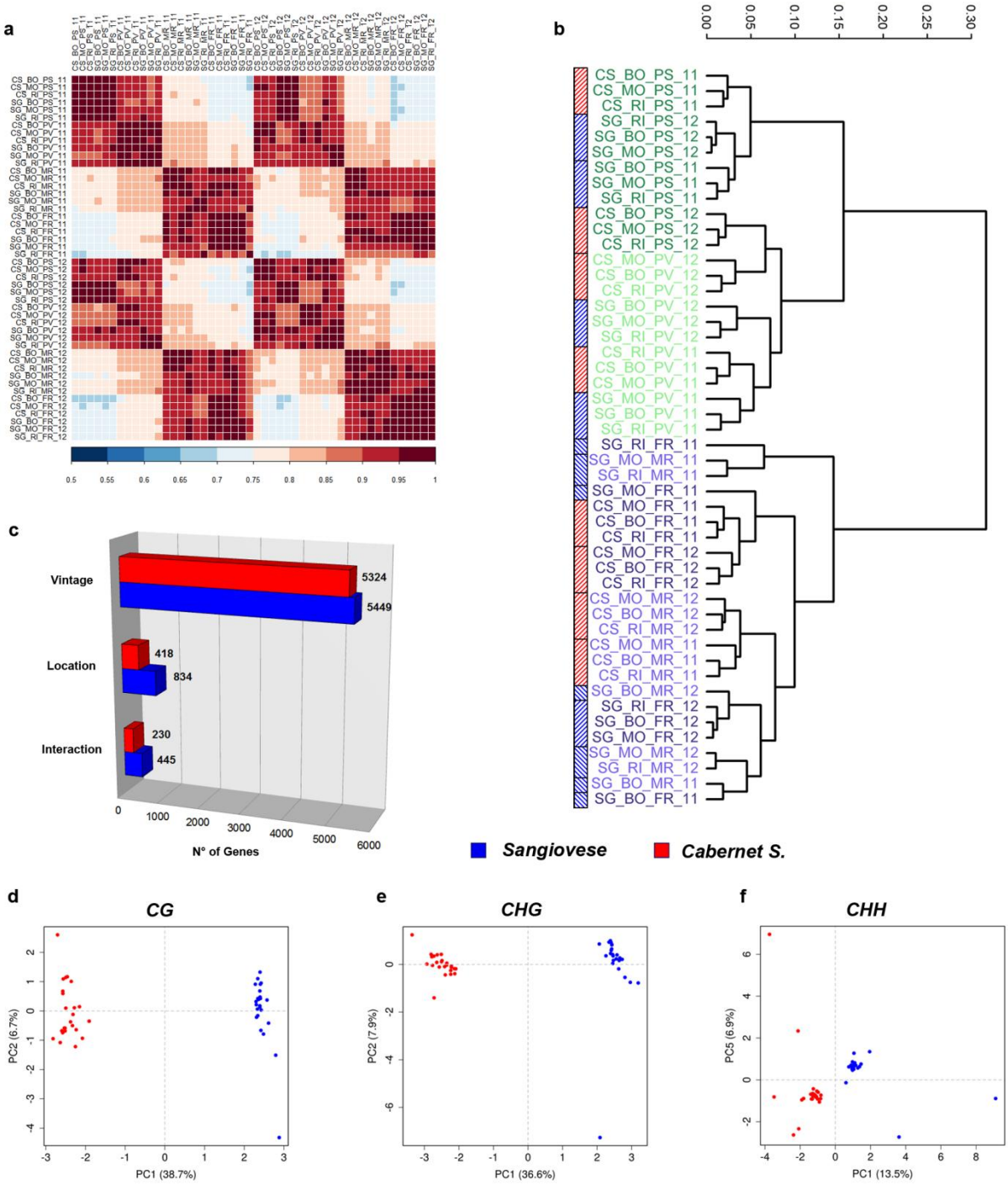
Figure 5. Characterization of GxE clusters of gene expression. **(A)** Venn diagram highlighting the GxE clusters, cultivar x vintage, cultivar x location, and cultivar x vintage x location. Dataset S7 provides a complete description of each cluster. **(B)** Gene Ontology (GO) analysis within the GxE cluster genes. The enriched GO biological processes were identified and listed according to their enrichment p-value ($p<0.05$). The total number of the GO category-related genes within the genes query is shown at the side of each bar. **(C-E)** Examples of GxE clusters of gene expression. **(C)** Cluster #295 (R*c* = 0.77) contains transcripts encoding the PRF disease-resistance protein, **(D)** Cluster #297 (R*c* = 0.89) contains members of the phenylalanine ammonia-lyase gene family, and **(E)** Cluster #300 (R*c* = 0.87) contains members of the terpene synthase gene family. See Dataset S3 for a complete description of the 300 clusters.

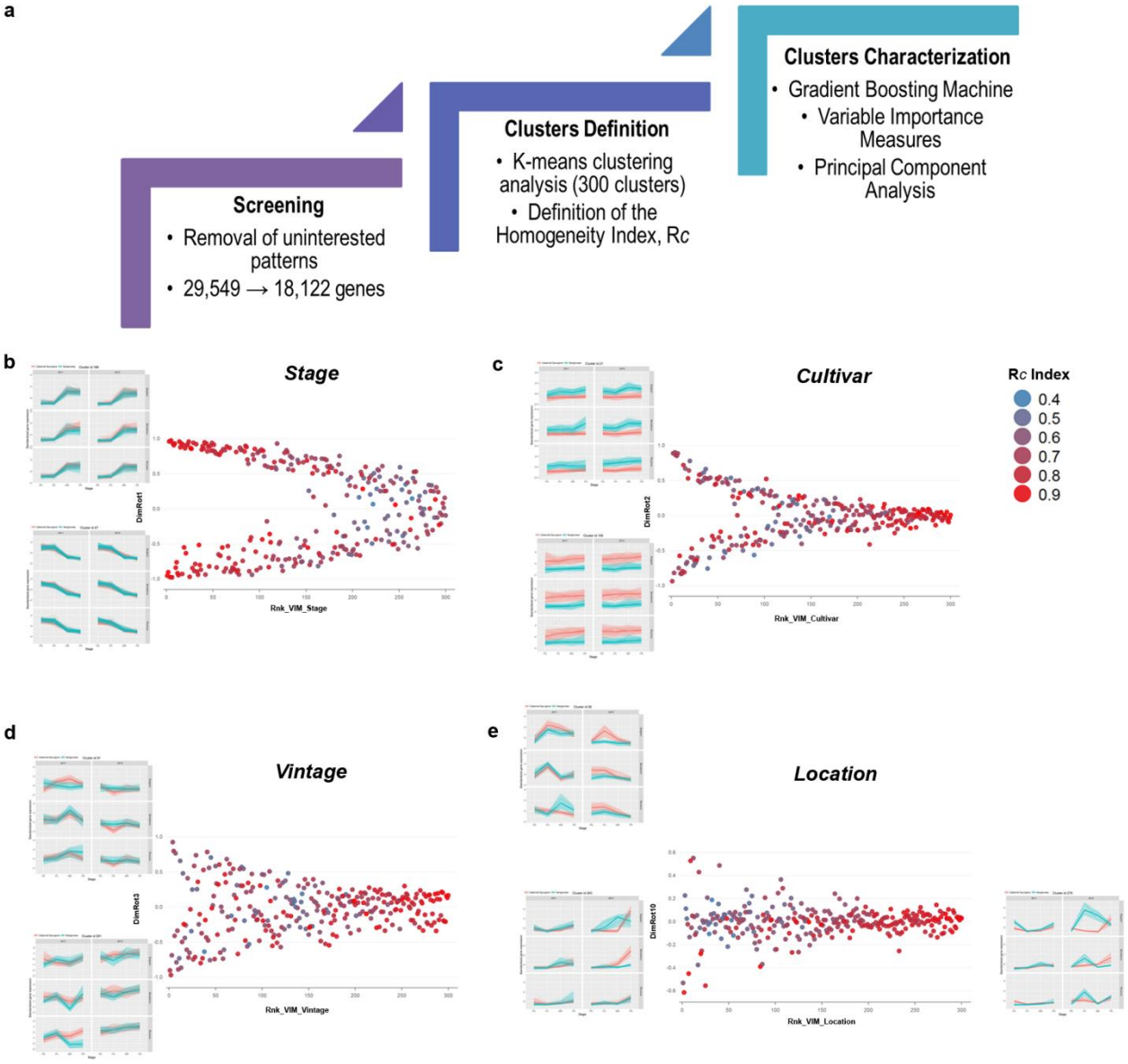
Figure 6. Genomic properties of variable-specific and variable-shared clusters of gene expression. **(A)** Haplotype sharing between Cabernet Sauvignon and Sangiovese across the 19 chromosomes. Black dots indicate the location of centromeric repeats. See Dataset S8 for more details. **(B)** Percentage of stage-dependent and cultivar-dependent genes sorted by their level of haplotype sharing. Asterisks indicate a significant difference (chi-square test, $p<0.01$) in the relative abundance of stage-dependent and cultivar-dependent genes and all genes, in regions of haplotype sharing 0, 1 and 2 (ns = not significant). **(C)** Percentage of modulated genes, sorted in classes based on their allelic state in Cabernet Sauvignon and Sangiovese, regardless of their

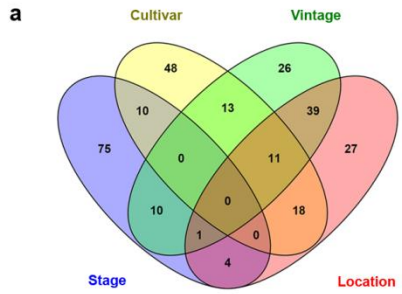
level of haplotype sharing. “Both Homozygous” = homozygous genes in both varieties; “Both Heterozygous” = heterozygous genes in both varieties; “One Homozygous, the Other Heterozygous” = genes homozygous in one variety and heterozygous in the other. See Dataset S9 for more details. Asterisks indicate a significant difference (chi-square test, $p<0.05$) in the relative abundance of each gene class in a specific cluster compared to all genes (ns = not significant).

Figure 7. Correlation between transcriptomic and climatic/agricultural data. **(A)** Correlation across the whole dataset. Left panel, correlation matrix (Spearman’s coefficient) prepared using the mean standardized expression value of each of the 300 clusters and climatic/agricultural data recorded during the whole experiment timespan (48 conditions). Central panel, DimRot1 heat map. Positive DimRot1 values indicate upregulation trends whereas negative DimRot1 values indicate downregulation trends (Figure 3b). Right panel, heat map of the VIM ranking positions per each variable. See Dataset S10 for more details. **(B-C)** Differences in transcriptomic plasticity between the two genotypes in the interaction with the environment. Subtraction matrices of the correlation matrices obtained for **(B)** the pre-veraison samples and **(C)** the post-veraison samples. Gray coloring shows where subtraction was not calculated (initial Spearman’s correlation value < 0.6 in either of the two genotypes). White coloring indicates subtraction value ≤|0.65|. Increasing green and purple intensity indicate subtraction values >|0.65| for pre-veraison and post-veraison matrixes, respectively. Daily mean (Td_5d), daily maximum (Tx_5d), daily minimum (Tm_5d) temperatures, thermal excursion [(Tx–Tm)_5d], Global Solar Radiation (GSR_5d) and rainfall (Rainfall_5d) were computed within the 5 days prior to each sampling date. GSR and rainfall were also computed on the whole timespan of the experiment (GSR_total and Rainfall_total). HWI, heat waves index; GDD_10C, growing degree days; SWC, soil water content; Pn, photosynthetic rate; E_T, transpiration rate; gs, stomatal conductance; HCA, hydroxycinnamic acids.



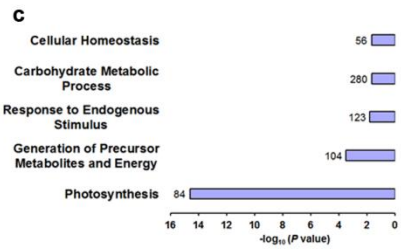


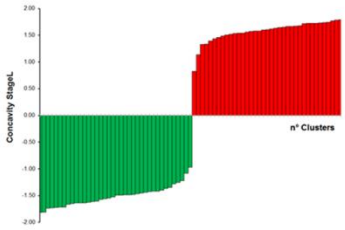


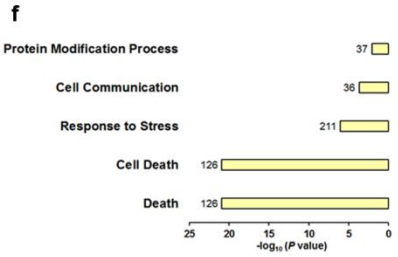


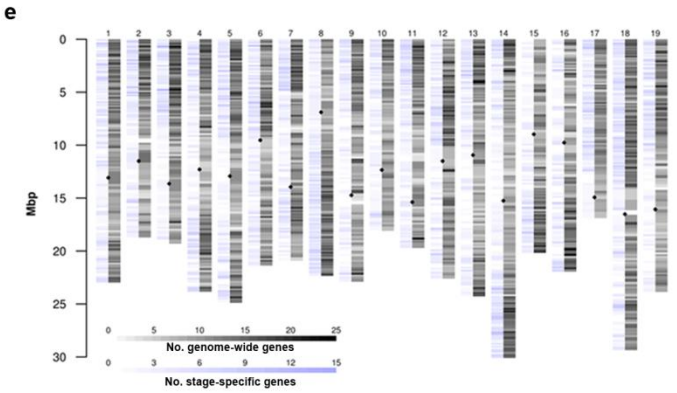
b

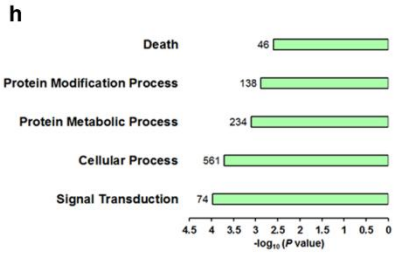
Variable Specific Clusters				
Variable	n° Clusters	Tot. Genes	Avg Genes/Cluster	R_c
Stage	75	6,793	90.57	0.82
Cultivar	48	2,648	55.17	0.72
Vintage	26	1,657	63.73	0.71
Location	27	1,183	43.81	0.63
Variable Shared Clusters				
Two-way Interactions				
Stage x Cultivar	10	621	62.10	0.81
Stage x Vintage	10	766	76.60	0.81
Stage x Location	4	232	58.00	0.74
Cultivar x Vintage	13	649	49.92	0.68
Cultivar x Location	18	635	35.28	0.60
Vintage x Location	39	1,478	37.90	0.64
Three-way Interactions				
Stage x Vintage x Location	1	61	61.00	0.81
Cultivar x Vintage x Location	11	434	39.45	0.61

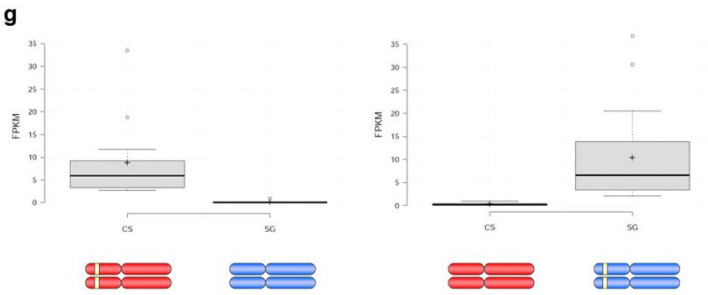


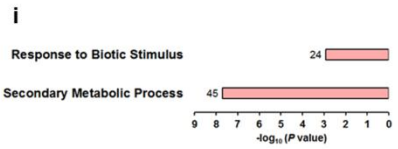


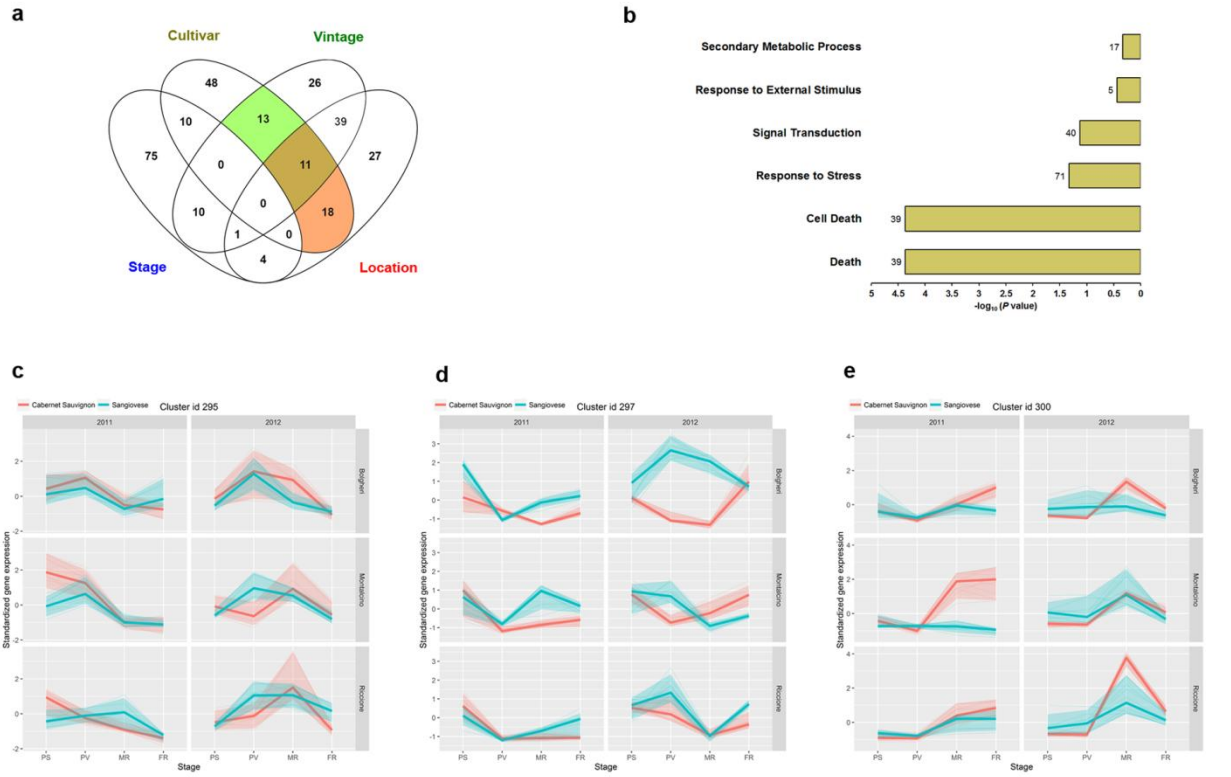












b

	Sharing 0	Sharing 1	Sharing 2	Total N° Genes
Total Genome	50.86	33.68	3.22	29,971
Stage	51.41 ns	***36.57*** *	3.31 ns	6,793
Cultivar	***58.80*** *	***27.61*** *	***1.32*** *	2,648

c

	Both Homozygous	Both Heterozygous	One Homozygous, the other Heterozygous	Total N° Genes
Total Genome	26.01	35.99	38.00	24,637
Stage	26.02 ns	34.91 ns	39.07 ns	6,056
Cultivar	***22.86*** *	36.65 ns	40.49 ns	2,139
Vintage	***30.40*** *	***32.62*** *	36.98 ns	1,398
Location	26.09 ns	36.85 ns	37.06 ns	966
GxE	24.91 ns	***32.76*** *	***42.33*** *	1,401