



**HAL**  
open science

## Genetic structure and first genome-wide insights into the adaptation of a wild relative of grapevine, *Vitis berlandieri*

Louis Blois, Marina de Miguel, Pierre-François Bert, Nabil Girollet, Nathalie Ollat, Bernadette Rubio, Vincent Segura, Kai Voss-Fels, Joachim Schmid, Elisa Marguerit

### ► To cite this version:

Louis Blois, Marina de Miguel, Pierre-François Bert, Nabil Girollet, Nathalie Ollat, et al.. Genetic structure and first genome-wide insights into the adaptation of a wild relative of grapevine, *Vitis berlandieri*. *Evolutionary Applications*, 2023, 16 (6), pp.1184-1200. 10.1111/eva.13566 . hal-04221476

**HAL Id: hal-04221476**

**<https://hal.inrae.fr/hal-04221476>**

Submitted on 18 Oct 2023



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

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



Distributed under a Creative Commons Attribution 4.0 International License

# Genetic structure and first genome-wide insights into the adaptation of a wild relative of grapevine, *Vitis berlandieri*

Louis Blois<sup>1,2</sup>  | Marina de Miguel<sup>1</sup>  | Pierre-François Bert<sup>1</sup> | Nabil Girollet<sup>1</sup> | Nathalie Ollat<sup>1</sup> | Bernadette Rubio<sup>1</sup> | Vincent Segura<sup>3</sup> | Kai P. Voss-Fels<sup>2</sup> | Joachim Schmid<sup>2</sup> | Elisa Marguerit<sup>1</sup>

<sup>1</sup>EGFV, Univ. Bordeaux, Bordeaux Sciences Agro, INRAE, ISVV, Villenave d'Ornon, France

<sup>2</sup>Department of Grapevine Breeding, Geisenheim University, Geisenheim, Germany

<sup>3</sup>AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France

## Correspondence

Louis Blois, EGFV, Univ. Bordeaux, Bordeaux Sciences Agro, INRAE, ISVV, Villenave d'Ornon F-33882, France.  
Email: [louis.blois@inrae.fr](mailto:louis.blois@inrae.fr)

## Funding information

Conseil Régional Aquitaine; Hessisches Ministerium für Wissenschaft und Kunst

## Abstract

In grafted plants, such as grapevine, increasing the diversity of rootstocks available to growers is an ideal strategy for helping plants to adapt to climate change. The rootstocks used for grapevine are hybrids of various American *Vitis*, including *V. berlandieri*. The rootstocks currently use in vineyards are derived from breeding programs involving very small numbers of parental individuals. We investigated the structure of a natural population of *V. berlandieri* and the association of genetic diversity with environmental variables. In this study, we collected seeds from 78 wild *V. berlandieri* plants in Texas after open fertilization. We genotyped 286 individuals to describe the structure of the population, and environmental information collected at the sampling site made it possible to perform genome–environment association analysis (GEA). *De novo long-read whole-genome sequencing was performed on V. berlandieri* and a STRUCTURE analysis was performed. We identified and filtered 104,378 SNPs. We found that there were two subpopulations associated with differences in elevation, temperature, and rainfall between sampling sites. GEA identified three QTL for elevation and 15 QTL for PCA coordinates based on environmental parameter variability. This original study is the first GEA study to be performed on a population of grapevines sampled in natural conditions. Our results shed new light on rootstock genetics and could open up possibilities for introducing greater diversity into genetic improvement programs for grapevine rootstocks.

## KEYWORDS

genome-wide association, genotyping by sequencing, grapevine, long reads, population genetics, rootstock, whole-genome sequencing

## 1 | INTRODUCTION

In the context of climate change, the resilience of plants in natural populations and the productivity of agronomic species are compromised (Parmesan & Yohe, 2003; Root et al., 2003). The increasing

frequency and intensity of threats such as drought, nutrient shortages, and the arrival of new pests and pathogens are introducing new challenges into plant breeding programs, which are increasingly called upon to develop new varieties able to overcome these environmental pressures.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd.

Natural selection has tested a much wider range of gene combinations under various environmental pressures than could ever be tested in plant breeding programs (Cortés & López-Hernández, 2021). For this reason, studies of the genetic basis of the adaptation of wild relatives of cultivated species to their native environments can provide useful genetic potential for incorporation into breeding programs (Condon et al., 2004; Vadez et al., 2014).

In recent years, improvements in sequencing technologies have made deeper explorations of the genetic basis of phenotype variability possible. Genome-wide quantitative genetic studies can identify markers associated with traits of interest. Following their identification and validation, these quantitative trait loci (QTL) can be used in marker-assisted selection (MAS; Tuberosa et al., 2007) to improve agronomic traits with a simple genetic architecture in crop plants; this approach has been used to increase yield in maize, rice, barley, and soybean (Francia et al., 2005). Genome-wide association studies (GWAS) are based on the combination of ancestral recombination events (Nordborg & Tavaré, 2002) and information about linkage disequilibrium (LD); they can be used to detect associations between allelic and phenotypic variation. This approach makes it possible to hone in on the positions of loci controlling traits of interest by making use of the large number of recombination events that have occurred over many generations in natural populations. The main limitation of GWAS has been the extent of LD, which can be affected by many factors, including structuring of the population, population size, and genetic drift (Remington et al., 2001), leading to high false discovery rates and inconsistent results. Population structure and kinship are two major confounding factors in the detection of genotype–phenotype associations (Kang et al., 2008, 2010; Santure & Garant, 2018; Zhang et al., 2010). However, it is possible to control for the false discovery rate in current statistical models without increasing computing time by considering kinship, structure, and LD (Huang et al., 2018).

A particular case of GWAS uses environmental variables instead of phenotypes to identify a link between genetics and environment (genome–environment association, GEA) in wild populations (Santure & Garant, 2018). This makes it possible to explore the genetic basis of adaptation to the environment, and to identify loci with allele frequencies correlated with climatic data (Bragg et al., 2015). GEA studies are complementary to GWAS, as they can reveal adaptive patterns that are difficult to detect with GWAS approaches, and can identify the major environmental forces behind natural selection (Rellstab et al., 2015). Approaches of this type have been applied to the wild relatives of several crops, including barley (Abebe et al., 2015), as a means of identifying putative adaptive loci and selecting gene pools adapted to specific environmental conditions. In perennial species, such as trees, GEA has been successfully used to identify populations displaying potential preadaptation to the predicted future climate (De La Torre et al., 2019; Pluess et al., 2016). It could ultimately be used in the development of genetic markers to assist breeding strategies and to facilitate the precise selection of new wild genotypes for inclusion in breeding programs (Cortés et al., 2022).

The genetic load of breeding populations has also been identified as one of the main challenges in the transition to next-generation breeding (Wallace et al., 2018). According to population genetics theory, most of the new mutations occurring in a population are neutral or slightly deleterious (Kimura, 1983; Ohta, 1973). Mutations with a strong deleterious effect should be rapidly eliminated by purifying selection. However, the efficacy of purifying selection for removing harmful alleles may be compromised in certain situations, such as demographic bottlenecks (González-Martínez et al., 2017; Peischl et al., 2013), and in the presence of Hill–Robertson interference (i.e., a phenomenon that links alleles with potentially different fitness values in regions of low recombination; Hill & Robertson, 1966). Decreases in the efficacy of purifying selection lead to the accumulation of deleterious mutations that may compromise the fitness of natural populations or the productivity and resilience of crop species. The identification, control, and repair of such mutations in major crop species is, therefore, crucial for the persistence of natural populations and for breeding programs.

Grapevine (*Vitis vinifera*) breeding programs are particularly challenging because this perennial plant species has been grafted onto rootstocks since the phylloxera crisis in the nineteenth century. *Vitis vinifera* is used as the scion, to maintain grape yield and quality, but other *Vitis* species resistant to phylloxera are used as rootstocks. These rootstocks are derived from hybridizations between American species (mostly *V. riparia*, *V. rupestris*, and *V. berlandieri*) and sometimes with *V. vinifera* (Galet, 1988). Their selection for use in vineyards is based on their tolerance to phylloxera, chlorosis and water-deficit tolerance, and the vigor conferred. Several studies have investigated genetic structure in grapevine (Aradhya et al., 2003; Arroyo-García et al., 2006; Cipriani et al., 2010; Frenkel et al., 2012; Grassi et al., 2003; Imazio et al., 2006; Laucou et al., 2011, 2018; Myles et al., 2010, 2011; Péros et al., 2011, 2015, 2021). Rootstock identification is traditionally based on ampelographic traits (Galet, 1956; Ravaz, 1902) and genetic studies have essentially been restricted to *V. vinifera*, with little effort devoted to grapevine rootstock genetics (Arroyo-García et al., 2006; Bianchi et al., 2020; de Andrés et al., 2007; Myles et al., 2011; Péros et al., 2011, 2015, 2021). New crosses are being performed to develop additional grapevine rootstocks, but mostly with the genotypes available in germplasm collections, and very little exploration of the genetic diversity existing in nature (Riaz et al., 2019). However, one recent study (Péros et al., 2021) based on SSR and SNP markers revealed a high level of genetic diversity in 421 genotypes of *V. aestivalis*, *V. cinerea* (var. *berlandieri* and *cinerea*), and *V. riparia*.

*Vitis berlandieri* is commonly used in crosses for the development of new grapevine rootstocks. Its hybrids perform well, but are difficult to use, mostly due to poor root emission after grafting (Boubals, 1966; Galet, 1988). However, the *V. berlandieri* genetic background is involved in hybridization to produce a number of widely used grapevine rootstocks, such as 110 Richter (*Vitis berlandieri* cv. Ressaygues no. 2 × *Vitis rupestris* cv. Martin), Fercal (Berlandieri Colombard no. 1 B × 31 Richter), Gravesac (161-49 Couderc × 3309 Couderc), and SO4 (*Vitis berlandieri* × *Vitis riparia*).

The *V. berlandieri* genotypes used for hybrid creation were selected rapidly at the start of the twentieth century on the basis of surveys performed across the USA. Unfortunately, this genetic background has never been explored more deeply, and the diversity and genetic architecture of traits of interest in *V. berlandieri* in the wild remains unknown. This species is endemic to the Edwards Plateau area in Texas (USA), a dry, chalky region (Schmid et al., 2009). In a previous study on American genetic backgrounds, *V. berlandieri*, which is considered to belong to the *V. cinerea* subgroup, was clearly separated from other groups (Péros et al., 2021). A GEA approach was carried out on 130 *Vitis* accessions including 22 *V. berlandieri* accessions to explore the SNP associations between bioclimatic variables and bacterial levels after infection with *Xylella fastidiosa*, the causative agent of Pierce's disease (Aguirre-Liguori et al., 2022; Morales-Cruz et al., 2021). Moreover, whole-genome sequencing studies were carried out with *V. riparia* (Girollet et al., 2019) and 21 *Vitis* accessions (Liang et al., 2019), but it remains a neglectable part of the grapevine studies more focused on *Vitis vinifera* diversity for scion-related traits (Tello & Ibañez, 2023). Thereby, the genetic background of wild *Vitis* remain poorly explored.

In this study, we addressed the following objectives: (i) characterization of the genetic structure of *V. berlandieri*; we generated genome-wide molecular markers from a de novo assembly of the *V. berlandieri* genome for this purpose; (ii) exploration of the genomic features of this species in terms of the extent of linkage disequilibrium and genetic diversity and their variation according to genetic structure; (iii) study of deleterious allele accumulation in the different subpopulations; and (iv) identification of genes potentially involved in the adaptation of this species to the environment.

## 2 | MATERIALS AND METHODS

### 2.1 | Plant material

In 2005, 78 wild female *V. berlandieri* plants were ampelographically identified in Edwards Plateau (transect of 40,000 km<sup>2</sup>; from N 31°23' W 100°2' to N 29°43' W 97°26') in Texas, USA. The coordinates and elevation of each sampling site were recorded. We harvested approximately 40,000 seeds from these plants after open fertilization, and about 5000 of these seeds were sown in a field at Geisenheim University, Germany. We selected 286 genotypes within this population on the basis of vigor, such that each initial "mother" plant was represented by four genotypes (half-sibs) to encompass the available genetic diversity.

### 2.2 | Reference genome

The reference genotype for this study was "*V. berlandieri* 10585" (NCBI, BioProject ID: PRJNA886625) from the collection of INRAE-Bordeaux, (Villenave d'Ornon, France). This genotype was selected

on the basis of its rooting capacity and use in other experiments. "*V. berlandieri* 10585" leaves were harvested from the INRAE Bordeaux grapevine collection (Villenave d'Ornon, France). Two young leaves with a width of about 5 cm were collected, frozen, and stored in a -80°C freezer for DNA extraction. DNA was extracted with the Tip 100 Qiagen Genomic kit, according to a slightly modified version of the manufacturer's protocol. We incubated 0.5 g of ground plant material with 9.5 mL of G2 buffer supplemented with 1% PVP-40, 19 µL RNase A, and 500 µL proteinase K for 3 h at 50°C for lysis. The lysate was subjected to tip filtration and the DNA was precipitated with isopropanol, centrifuged for 15 min at 5000g, washed with 70% ethanol, and resuspended in 50 µL TE buffer. The quality and molecular weight of the DNA isolated were checked. An A<sub>260</sub>/A<sub>280</sub> ratio between 1.8 and 2.0 and an A<sub>260</sub>/A<sub>230</sub> ratio between 2.0 and 2.2 were obtained, and an Agilent Genomic DNA Screentape analysis was performed. We used 10 µg of high-quality DNA for sequencing. Samples were sequenced with Single-Molecule Real-Time PACBIO SEQUEL II HIFI long reads at the INRAE Clermont-Ferrand GENTYANE platform (France).

DNA consensus call sequences obtained in BAM format were converted to Fastq format with the bam2fastq tool from the SMRTLink v11.0 PACBIO library. The HIFI sequencing DNA quality was checked with FastQC version 0.11.7. Paternal and maternal kmers were identified with yak-0.1 software, using the parental reads. The outputs were then used in hifiasm-v0.15.5 to bin long reads and to assemble the two haplotypes. For each haplotype, contigs were aligned with PN40024.v4 with minimap2 version 2.17. The best contig alignments were used to build an AGP file, and each pseudomolecule was then reconstructed. We refined the pseudomolecules by repeating the process, beginning with an alignment of each haplotype against the other, previously reconstructed haplotype. The embryophyta\_odb10 lineage package from BUSCO 5.3.1 software was used in genome mode to estimate the completeness of all assemblies.

### 2.3 | Genotyping by sequencing

Leaves were sampled from all 286 genotypes at Geisenheim University, Germany. Two leaf discs of 1.5 cm diameter were sampled from each genotype and placed on ice. The leaf discs were then frozen in liquid nitrogen and freeze-dried in a Martin Christ, Beta 2-8LD freeze-dryer. DNA was extracted from all lyophilized samples as described by Cormier et al. (2019) in Corning/Costar deep 96-well 1.1 mL plates. Libraries were prepared at UMR AGAP, CIRAD (Montpellier, France) as described by Elshire et al. (2011). Based on our results, the amount of DNA was normalized to 50 ng/mL. We prepared 96-plex GBS banks with the restriction enzyme ApeKI. Illumina Hiseq 4000 short-read 2 × 150 bp sequencing was performed by Genewiz. Three 96-plex plates were used. The raw data are available on NCBI, BioProject ID: PRJNA886619.

## 2.4 | SNP calling

GBS data were processed with the Genotoul cluster, Toulouse, France. Reads were demultiplexed with a script available from <https://github.com/timflutre> (demultiplex.py) and cleaned with Cutadapt (Martin, 2011) with filters "-a AGATCGGAAGAGCGGT CAGCAGGAATGCCGAG -A GAGATCGGAAGAGCGTCTGTAGG GAAAGAGTGT -G CTCGGCATTCTGCTGAACCGCTTCCGAT CT -g AACTCTTTCCCTACACGACGCTCTTCCGATCT -u 7 -U 7 -m 17." Sequences were aligned with the *V. berlandieri* reference genome obtained in this study. VCFs were joined with GATK tools (McKenna et al., 2010), and 3,294,984 SNPs were obtained. SNPs were filtered, with the rejection of SNPs with a quality depth < 2.0, Fisher strand value > 60.0, MQ < 40.0, MRankSum < -12.5, and ReadPosRankSum < - 8.0. These filters were applied one by one, as recommended in the GATK support documentation. In total, 3,294,747 SNPs were conserved and individuals with more than 80% missing data were filtered out ( $n=281$  genotypes retained). VCFtools (Danecek et al., 2011) was used for filtering based on the following criteria: minimum depth of 3, maximum of 50% missing data, minor allele frequency of 0.05, and a minimum mean depth of 5. In total, 104,378 SNPs were conserved. We considered 281 genotypes with less than 60% missing data.

## 2.5 | Population structure

Two methods were used for the analysis of population structure. We first ran STRUCTURE v2.3.4 (Pritchard et al., 2000) on all 104,378 SNPs. The optimal number of subpopulations was determined as previously described (Evanno et al., 2005). One to 10 populations were allowed, with a burn-in period of 20,000 and a Markov Chain-Monte Carlo (MCMC) iteration number of 20,000 without prior knowledge of population affinities and three runs. The optimal number of populations was found to be two ( $K=2$ ). A new run was then performed, with a burn-in period of 100,000 and 100,000 MCMC iterations. Genotypes with a membership proportion greater than 0.8 for a population were attributed to the population concerned. Genotypes not attributed to a particular population were considered to be admixed. The STRUCTURE results were then compared with those of a  $k$ -means clustering method described elsewhere (Voss-Fels et al., 2015). A genetic matrix distance was calculated with Roger's distance (RD). Clusters were identified by the unweighted pair-group with arithmetic mean (UPGMA) method, based on Roger's genetic distance. Genotypes were then attributed to a cluster by the  $k$ -means clustering method, with the Hartigan and Wong algorithm (1979). The optimal number of clusters was determined by plotting cluster numbers from 1 to 15 against the corresponding within-cluster sum of squares (Voss-Fels et al., 2015). This calculation was performed 20 times and the mean value of each run was reported. The optimal number of clusters was found, by eye, to be  $K=5$ , as beyond this value, increasing the number of clusters did not significantly decrease the within-cluster sum of squares. The results

were visualized by plotting PCA results constructed with the first four principal components according to marker information. The number of clusters in the PCA was chosen based on the optimal  $K$  value from STRUCTURE analysis considering admixed individuals as a subpopulation ( $K=3$ ) and from the visualization of within-cluster sums of squares ( $K=5$ ).

We tested for isolation by distance by calculating Nei's distance matrix between genotypes with the *adegenet* package in R. As there were several genotypes originating from each mother plant, the mean genetic distance was calculated between mother plants so as to obtain a single value for the genetic distance between two mother plants. A distance matrix was constructed from the GPS coordinates of the mother plants with the *sp* package in R. A Mantel correlation test was performed by the Spearman method, with 9999 permutations.

## 2.6 | Linkage disequilibrium and genetic load

Linkage disequilibrium decay was estimated as the physical distance at which  $r^2$  reached a value of 0.2, as previously described (Hill & Weir, 1988). Intrapopulation  $F_{ST}$  was calculated by STRUCTURE and interpopulation ( $K=2$ )  $F_{ST}$  was calculated with VCFtools (--weir-fst-pop option), using the default parameters. The *V. berlandieri* reference genome was annotated according to the Pinot noir reference genome (12X.v2; Canaguier et al., 2017), with liftoff1.6.1. In total, 39,250 of the 42,413 genes from the Pinot noir reference genome (92.5%) were found in *V. berlandieri*. The annotated *V. berlandieri* genome was added to the program as an additional reference genome. SnpEff (Cingolani et al., 2012) was run with default parameters, and two items of information were considered in our analyses: the predicted impact of the SNP on the DNA sequence and the effect of the SNP according to its location in the genome (see Table S2 for a detailed list of effects). Impacts were classified into four categories: "high" (e.g., loss of function of the protein), "moderate" (e.g., modification of protein efficacy), "low" (probably no impact on protein), and "modifier" (mostly noncoding variants or variants for which there was no evidence of impact). Kimura's neutral theory of evolution (Kimura, 1968) suggests that most mutations are deleterious or neutral. We therefore calculated the genetic load as follows:

$$P_{del} = \frac{\text{Number of minor alleles for "high" impact SNP}}{\text{Number of alleles for SNP with "moderate" and "high" impact} \times 2} \quad (1)$$

This Equation (1) considers " $P_{del}$ " as the proportion of deleterious alleles for one individual. SNPs with missing data were not considered for each individual. The accumulation of deleterious alleles in each population was evaluated as the mean  $P_{del}$  of the individuals assigned to each population according to STRUCTURE analysis divided by the number of deleterious alleles for all individuals from each population. The accumulation of deleterious alleles was compared

between subpopulations by a Tukey test. For each SNP, snpEff could propose several impacts and effects. In such cases, we retained the first effect/impact proposed by the program, which was the most deleterious.

## 2.7 | Genome-wide association with environmental variables

Genome–environment association (GEA) analysis was performed with the elevation of the mother plants reported during sample and additional environmental information extracted from the TerraClimate platform (Abatzoglou et al., 2018) for each of the mother plant coordinates. The following environmental parameters were obtained for the 1991–2020 period, at a resolution of 4 km:

- Growing season temperature (GST, from April to October, Jones, 2006) and growing season temperature during the vegetative period (GST49, between April and September). The mean daily average temperature is calculated over the period concerned. This parameter affects earliness and grape quality (Jones, 2006).
- Growing season rainfall (GSR from April to October, Bois et al., 2017) and growing season rainfall during the vegetative period (GSR49, from April to September). This parameter is calculated as the cumulative amount of rainfall (mm) over the period concerned. It provides a rough estimate of the amount of water available to the plant during the corresponding period.
- Springtime rainfall (RRSPR), which is essentially GSR for the period between April and July. It can be used as an indicator of biotic pressure during the first few months of vegetative growth (Bois et al., 2017).
- Branas hydrothermal index (HYB; Branas et al., 1946), which evaluates the risk of grapevine exposure to disease.
- Winter freeze risk index (WFR; Bois et al., 2014) is the mean minimum temperature in January. If this temperature is  $>4^{\circ}\text{C}$ , the risk is considered to be low, whereas the risk is considered high if it is  $<-11^{\circ}\text{C}$ .
- Spring frost risk index (SFR; Bois et al., 2014) is the mean minimum temperature in April. The risk is considered low if this temperature is  $>12^{\circ}\text{C}$  and high if it is  $<0^{\circ}\text{C}$ .
- Heat stress index (HST; Bois et al., 2014) is the mean maximum temperature in July. The risk is considered to be low if this temperature is  $<25^{\circ}\text{C}$  and high if it is  $>30^{\circ}\text{C}$ .
- The Huglin index (HI; Tonietto & Carbonneau, 2004) provides information about the climate of the region. It combines mean air temperature ( $T$ ,  $^{\circ}\text{C}$ ), the maximum air temperature ( $T_x$ ,  $^{\circ}\text{C}$ ), and day length coefficient ( $d$ ) according to latitude, between April and September. This parameter may reflect the sugar-producing potential of the plant during the vegetative period in a given context.

$$Hi = \sum_{01/04}^{30/09} \frac{(T - 10) + (T_x - 10)}{2} d$$

- Cool night index (CI, Tonietto & Carbonneau, 2004) is the mean daily minimum air temperature in September. It provides qualitative information about the potential of wine-producing regions based principally on the production of secondary metabolites (polyphenols, aromas).
- Dryness index (DI, Tonietto & Carbonneau, 2004) is an indicator of potential soil water availability according to the dryness of a climatic region. It is adapted from the soil index of Riou (Riou et al., 1994) and it affects ripening and wine quality (Carbonneau, 1998).

The Pearson coefficients of correlation between these environmental variables were explored. GST and GST49 were highly correlated, as were GSR and GSR49 (coefficients of 0.99 for both). We therefore included only GST49 and GSR49 in the PCA calculation. A GEA analysis was performed for elevation, and then for the first two principal components of the PCA performed with TerraClimate environmental parameters. Elevation was analyzed separately from the other environmental parameters because it was measured during the sampling campaign, whereas the other environmental variables were obtained by climatic interpolation.

We performed GEA with the BLINK model in GAPIT with default settings. Before running GWAS, the SNPs dataset has been imputed with the snp.impute function available in GAPIT using the “Major” argument in order to impute missing data by the major allele of each SNP. Because the imputation procedure led to changes in the allele frequency at each position, mainly for SNPs with a large number of missing data, a new  $\text{MAF} > 0.05$  filtration was applied excluding 15,574 SNPs. The BLINK model was used because of the ease with which false discovery rate can be controlled in this model (Huang et al., 2018). The information about population structure obtained from STRUCTURE ( $K=2$ ) was used as a covariate. In this model, kinship was derived from pseudo-QTN information. Bonferroni correction was applied to the calculated  $p$ -values. The significance thresholds were, thus, set at  $0.05/n$  and  $0.01/n$  where “ $n$ ” is the number of markers used. The GEA analysis was performed on the 281 genotypes with 88,804 SNPs after the recalculation of minor allele frequency in GAPIT.

The genes attributed to each marker by snpEff were used to identify the related protein families with the UniProtKB (The UniProt Consortium, 2021) database. If several results were available, only one was retained. If the corresponding protein was unknown, we used the molecular function attributed by IEA:InterPro (Camon et al., 2005) or the manual assertion based on the work of Gaudet et al. (2011).

## 3 | RESULTS

The long-read sequencing of *V. berlandieri* resulted in 1,620,094 reads  $>1$  kb and a coverage of 50 $\times$ . The N50 of reads was 15.6 kb and L50 was 680,049, indicating that 680,049 reads in which the smallest size was 15.6 kb were enough to obtain half of the 25,000,009,766

bases sequenced. In addition, 91% of reads had a length between 10 and 20kb. Then, 2789 contigs were obtained with a N50 of 22.26Mb and L50 of 9. When comparing with the *V. riparia* genome (Giollet et al., 2019), the *V. berlandieri* genome was well represented with 29 contigs. In the final version, 10 chromosomes presented a single contig, eight chromosomes had two contigs, and one chromosome was divided into three contigs. In total, 39,250 of the 42,413 genes from the Pinot noir reference genome (92.5%) were found in *V. berlandieri*. This high-quality assembly resulted in a highly reliable reference genome for these analyses.

### 3.1 | Population structure

The genotyping by sequencing of the *V. berlandieri* population allowed us to get 104,378 relevant SNPs with an average of 5500 SNPs per chromosome (Figure 1) and a mean of 0.2 SNPs per kb. From the marker set, 787 SNPs were detected with a high impact by snpEff, 17,388 SNPs with a low effect, and 18,215 with a moderate effect (Table S1). According to the results obtained with snpEff, most of SNPs were upstream\_gene\_variant (23,002), missense\_variant (18,102), synonymous\_variant (15,937), downstream\_gene\_variant (15,028), intron\_variant (14,022), or intergenic\_variant (10,622; Table S2). The optimal number of subpopulations determined by STRUCTURE was  $K=2$  according to the method of Evanno et al. (2005) (burn-in=20,000 and no. of MCMC iterations=20,000). The two subpopulations were explored in greater depth (burn-in=100,000 and no. of MCMC iterations=100,000; Figure 2a); subpopulation 1 was found to contain 100 genotypes, subpopulation 2 contained 63 genotypes, and the admixed subpopulation contained 118 genotypes.

The  $k$ -means clustering method suggested that there were more subdivisions within the population, resulting in at least five subpopulations, as suggested by the sum of squares curve (Figure S1). The subpopulations were also explored by PCA, with the genotypes colored according to the groups previously determined with STRUCTURE,

as described by Evanno et al. (2005) with the admixed group considered as a subpopulation,  $K=3$ , (Figure 2b) and according to the results obtained with the  $k$ -means clustering method (Hartigan & Wong, 1979;  $K=5$ , Figure S2). The PCA (Figure 2b) revealed a clear subdivision into groups for  $K=3$ , but the proportions of the variance explained by principal components 1, 2, and 3 were very low, at 3%, 2%, and 1%, respectively. For  $K=5$  (Figure S2), a clear separation of the groups was observed for dimensions 1/2 and 1/3. However, when dimensions 2/3 or 1/4 were used, group 3 was the only group that could be distinguished clearly. The IBD analysis revealed a correlation between genetic distance and physical distance between sampling points (Spearman's  $\rho=0.33$  and  $p$ -value= $10^{-04}$ ).

### 3.2 | Linkage disequilibrium and genetic load

The extent of LD decay ( $r^2 < 0.2$ ) was calculated per chromosome; it ranged from 307bp on chromosome 16 (Figure S3a) to 8kb on chromosome 6 (Figure S3b) with a mean value of 2.2kb (method from Hill & Weir, 1988). SNPs with a predicted high impact accounted for 0.75% of SNPs, and these SNPs were the least represented group. Similar proportions of SNPs were predicted to have moderate and low impacts (ca. 17%; Table S1). SNPs classified as modifiers were the most abundant (65%). The distribution of each category of SNPs was similar between chromosomes. The number of SNPs for each estimated effect is detailed in Table S2. There were no significant differences in the accumulation of deleterious alleles between subpopulations (Figure S4).

### 3.3 | Genome–environment association

The structure of the population followed a geographic pattern (Figure 3). The elevation of the sampling sites differed considerably (Figure S5) ranging from 722 to 2295m. The mean elevation of each subpopulation made it possible to separate the groups easily

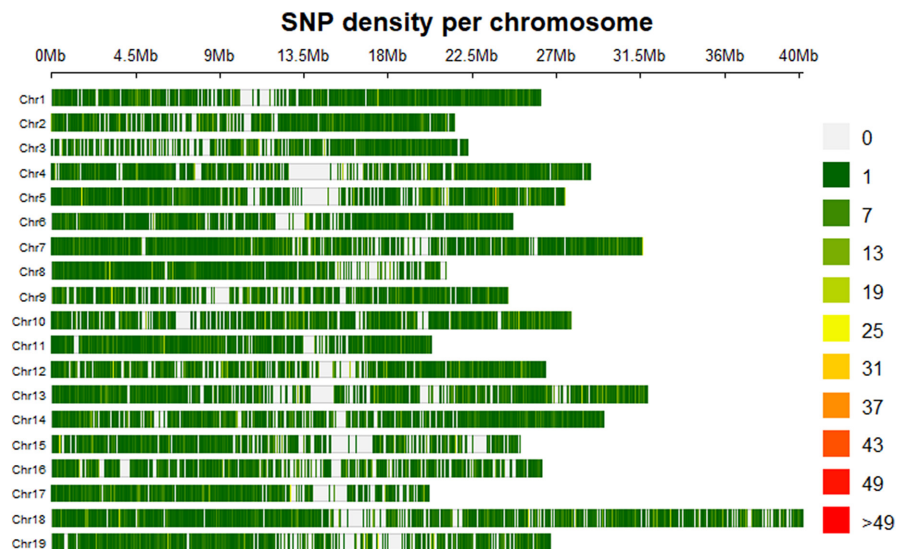
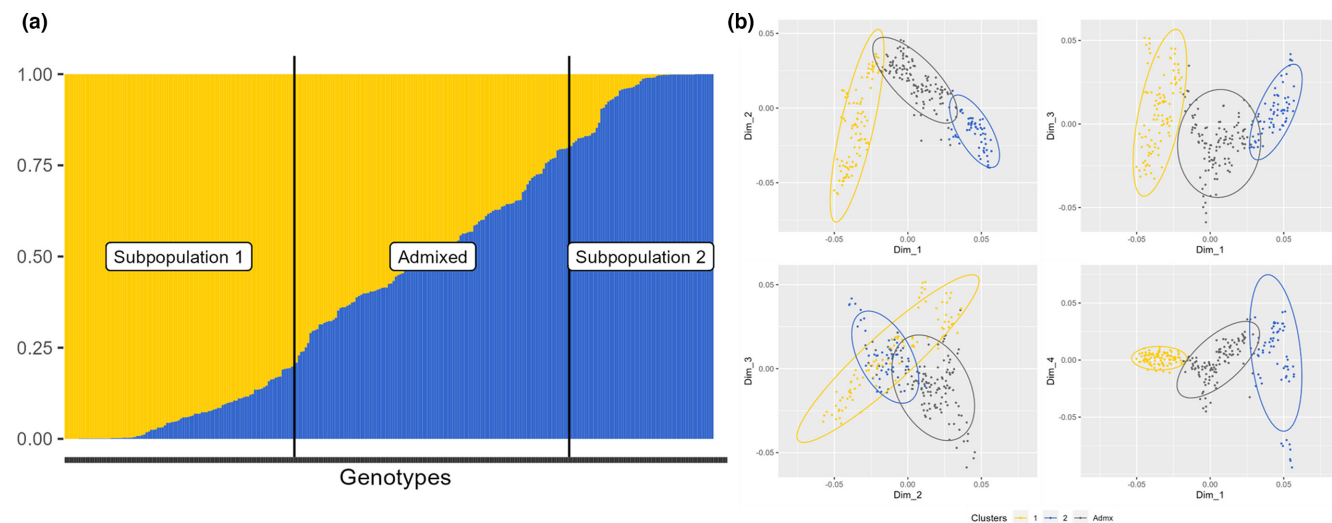


FIGURE 1 SNP density per kb obtained by GBS for each chromosome of the *Vitis berlandieri* genome.



**FIGURE 2** Population structure results from (a) STRUCTURE  $K=2$ . The proportion of each genotype found in the two populations is shown. Whenever a genotype has a probability of belonging to a population of 80% or more, the genotype is attributed to that population, each subpopulation being separated by a vertical black line. (b) PCA based on SNP information, the first four principal components are represented and all genotypes are plotted and colored according to STRUCTURE groups; PC 1–4 explained 4%, 2%, 2%, and 1% of the variability, respectively.

(Figure 4). Three QTL for elevation were identified on chromosomes 2 ( $p$ -value =  $9.60 \times 10^{-08}$ ), 7 ( $p$ -value =  $4.24 \times 10^{-08}$ ), and 15 ( $p$ -value =  $4.61 \times 10^{-10}$ ; Figure 5). Considerable variation was observed for environmental variables, with GSR ranging from 395 to 581 mm, HYB from 6932 to 9326 and DI ranging from  $-139$  to  $-29$  (Figure S6).

Subpopulation 1 was located in an area that was warmer and damper than the areas occupied by the other two subpopulations, with a higher GST49, GSR49, RRSPR, HST, WFR, and SFR. The HI value revealed a *very warm climate* ( $>3000$ ) for all subpopulations and was significantly higher for subpopulation 1 than for the other populations (Figure S6j); the CI value ( $>18$ ) indicated *warm nighttime* conditions for subpopulation 1.

Subpopulation 2 was located in a region that was drier and cooler than that occupied by subpopulation 1, with a moderate risk of frost during the spring ( $<12^\circ\text{C}$ ; Figure S6h) and temperate nights ( $14$ – $18^\circ\text{C}$ ; Figure S6k). Based on DI, the three subpopulations were considered to come from a *moderately dry* area, with a few genotypes from subpopulation 2 and the admixed subpopulation classified as coming from a *very dry* area, significantly drier than that for subpopulation 2 as a whole (Figure S6l).

As correlations were found between environmental parameters, PCA was performed to capture environmental variability in an integrated manner (Figure 6). The first two principal components (PC) explained 94% of the variance. The PC1 was principally explained by the parameters SFR and WFR which were related to frost risk during winter and spring. The PC2 was explained by HSI, HI, and DI which were related to high temperature and water availability. An analysis of the genome–environment association for the PC1 trait identified eight significant QTL on chromosomes 1, 4, 5, 6, 7 (2), and 19 (Figure 7). For PC2, seven significant QTL were detected on chromosomes 7 (2), 9 (2), 10, 14, and 18 (Figure 8). The significant QTL

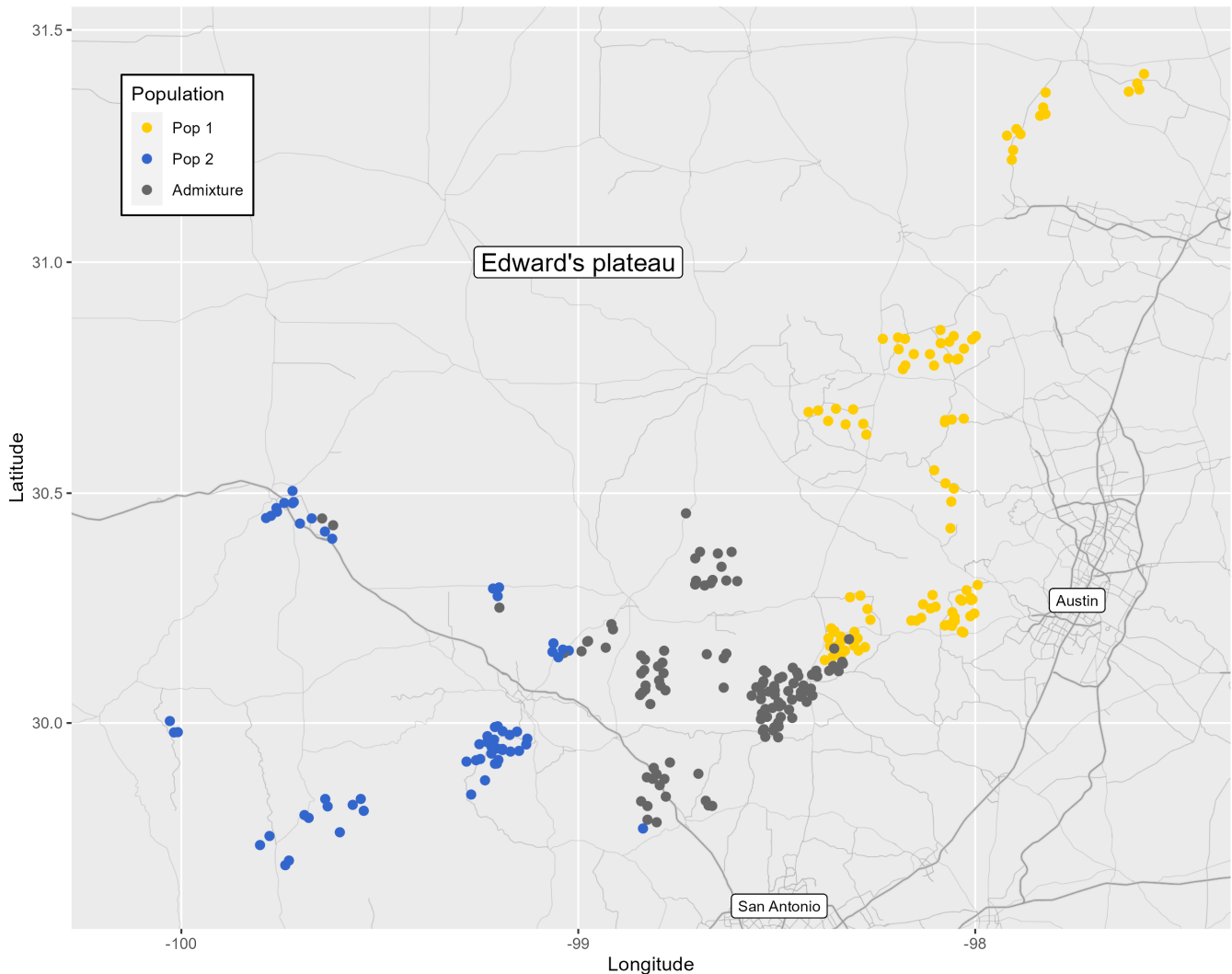
corresponded to SNPs classified as *modifiers*, or with *low* or *moderate* impact, and the genes concerned had basic molecular functions potentially involved in various biological pathways (Table 1). For each marker, the effect of each allele in the homozygous and heterozygous states have been explored (Figure S12).

## 4 | DISCUSSION

De novo whole-genome sequencing of *V. berlandieri* made it possible to explore the genetic structure of a wild *V. berlandieri* population with genome-wide molecular markers. We identified two distinct subgroups that could be explained by isolation by distance. However, no significant differences in genetic diversity or genetic load were found between subpopulations. Using geographic coordinates and climatic data from meteorological stations, we characterized the environments from which the subpopulations were collected and identified several QTL associated with environmental conditions and agronomic indices of climatic conditions.

The complexity of the plant genome renders short-read sequencing highly challenging, due to the large numbers of repetitive sequences and heterozygosity (Schatz et al., 2012). We used long-read sequencing, which overcomes these problems, in this study. The low level of LD in wild *Vitis* species is well known (Liang et al., 2019; Marrano et al., 2018; Myles et al., 2010, 2011; Péros et al., 2021). Our results confirm the rapid LD decay (about 2 kb) reported in previous studies. This rapid LD decay drives the retention of as many SNPs as possible, making it possible to capture a maximum of the genetic variability in the population (Flutre et al., 2019; Myles et al., 2011; Remington et al., 2001). We obtained a set of 104,378 SNPs, which we used for fine genetic association studies. We ensured that the





**FIGURE 3** Geographic position of the sites from which plants were sampled in Texas (the genotypes from the same mother plant have been jittered to facilitate observation).

maximum level of genetic diversity was retained, by including genotypes with up to 60% missing data in our analyses.

#### 4.1 | Genetic structure

Wild *Vitis* species are characterized by considerable genetic diversity (This et al., 2006). The genetic diversity and the relationships between these species and varieties has already been investigated, as interspecific hybrids and rootstocks have been shown to display greater genetic diversity than varieties of *V. vinifera* (de Andrés et al., 2007; Laucou et al., 2011). The wild grapevines (*V. vinifera* ssp. *sylvestris*) of the Anatolia region, which is considered to be the center of origin of grapevine (McGovern, 2003), are highly diverse (Ekhaia et al., 2014; Ergül et al., 2011). However, the genetic diversity of wild grapevines in the European region is lower (Di Vecchi-Staraz et al., 2009; Lopes et al., 2009), probably mirroring the human footprint on these populations. The genetic diversity of *Vitis* has been also explored in the Americas, revealing different

groups in South America (Martinez et al., 2006) and North America (Péros et al., 2021), and considerable diversity between species. However, genetic diversity and genetic structure within *Vitis* species has been little explored. We addressed this issue here, by studying the population structure and genetic diversity in *V. berlandieri*, a species of considerable interest for the breeding of grapevine rootstocks.

The population structure detected in our study could be explained geographically. The two subpopulations were from physically close locations and were poorly differentiated ( $F_{ST}=0.036$  and  $0.060$  for subpopulations 1 and 2, respectively, in STRUCTURE analysis), probably due to the small size of the geographic area used for sampling. The significant IBD could explain the low  $F_{ST}$  observed in each subpopulation, as it could indicate existing gene flow between subpopulations. Moreover, the  $F_{ST}$  between the two subpopulations was  $0.032$ . A low level of differentiation at regional level has also been observed in other perennial species, such as *Fagus sylvatica* (Buiteveld et al., 2007; Pluess et al., 2016), *Prunus sibirica* (Wang et al., 2014), and *Malus sieversii* (Richards et al., 2009).

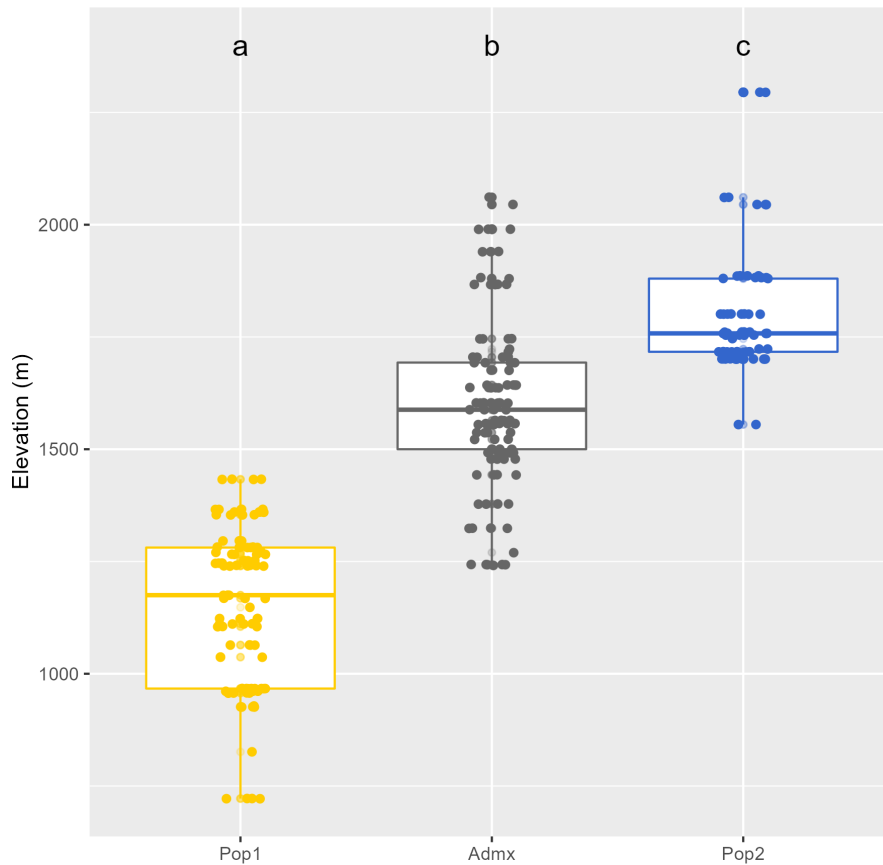


FIGURE 4 Boxplot of the elevations of the various subpopulations identified by STRUCTURE.

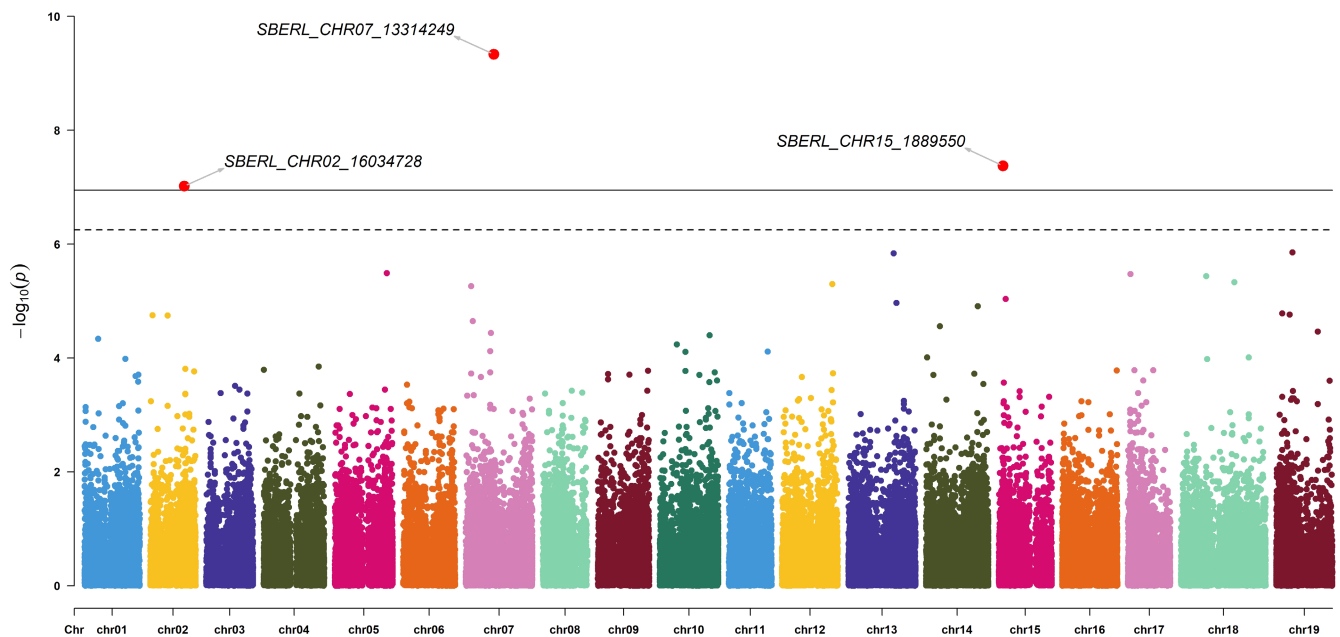
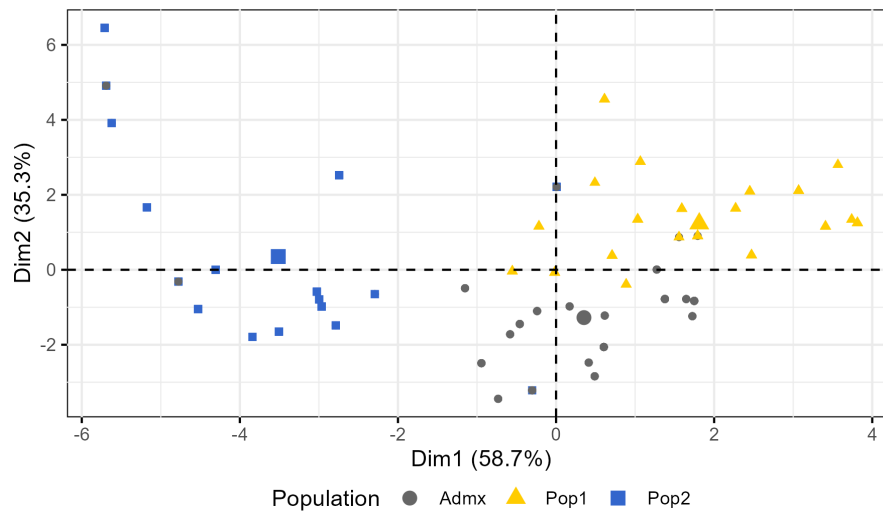


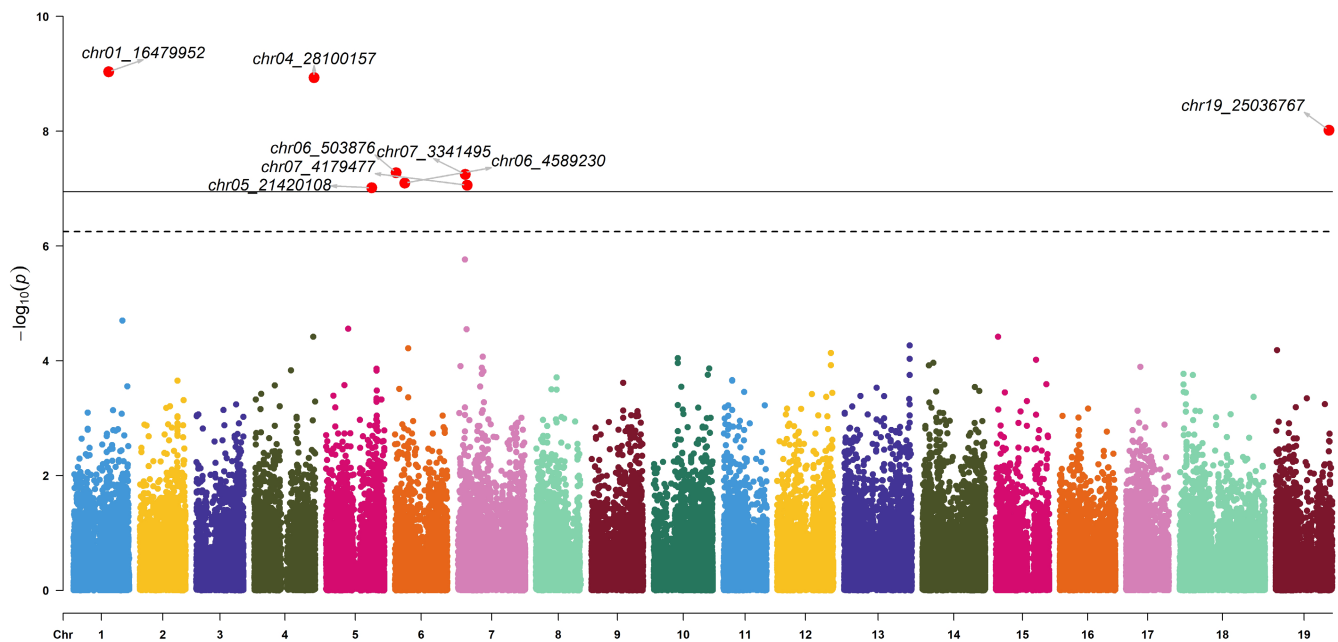
FIGURE 5 Manhattan plot for SNP associations with elevation. The thresholds were calculated by the Bonferroni method ( $\alpha/\text{NSNP}$ ) for  $\alpha=0.05$  (dotted line) and 0.01 (solid line). Significant signals are highlighted in red, and the corresponding QQ plot is presented in Figure S9.

This poor differentiation between subpopulations reveals the minor impact of this genetic structure, reducing the risk of false-positive detection in GWAS models (Santure & Garant, 2018). The sampling area in our study was restricted to the region in

*V. berlandieri* is endemic, but this species extends over a larger area (southern New Mexico, south-western Texas and northern Mexico; Galet, 1988) with different soil and climatic conditions. Greater diversity might, therefore, be expected for a larger



**FIGURE 6** PCA of the environmental parameters accessed via the TerraClimate platform (Abatzoglou et al., 2018). The following variables were extracted: growing season temperature between April and September (GST49, GST, Jones, 2006), growing season rainfall between April and September (GSR49, GSR, Bois et al., 2017), springtime rainfall (RRSPR, from April to July), Branas hydrothermal index (HYB, Branas et al., 1946), winter cold damage index (WFR, Bois et al., 2014) spring frost index (SFR, Bois et al., 2014), heat stress index (HST, Bois et al., 2014), Huglin index (HI, Tonietto & Carbonneau, 2004), cool night index (CI, Tonietto & Carbonneau, 2004), and dryness index (DI, Tonietto & Carbonneau, 2004). The PC1 is mainly explained by SFR and WFR. The PC2 is mainly explained by HSI, HI, and DI. Each group identified by STRUCTURE is indicated.

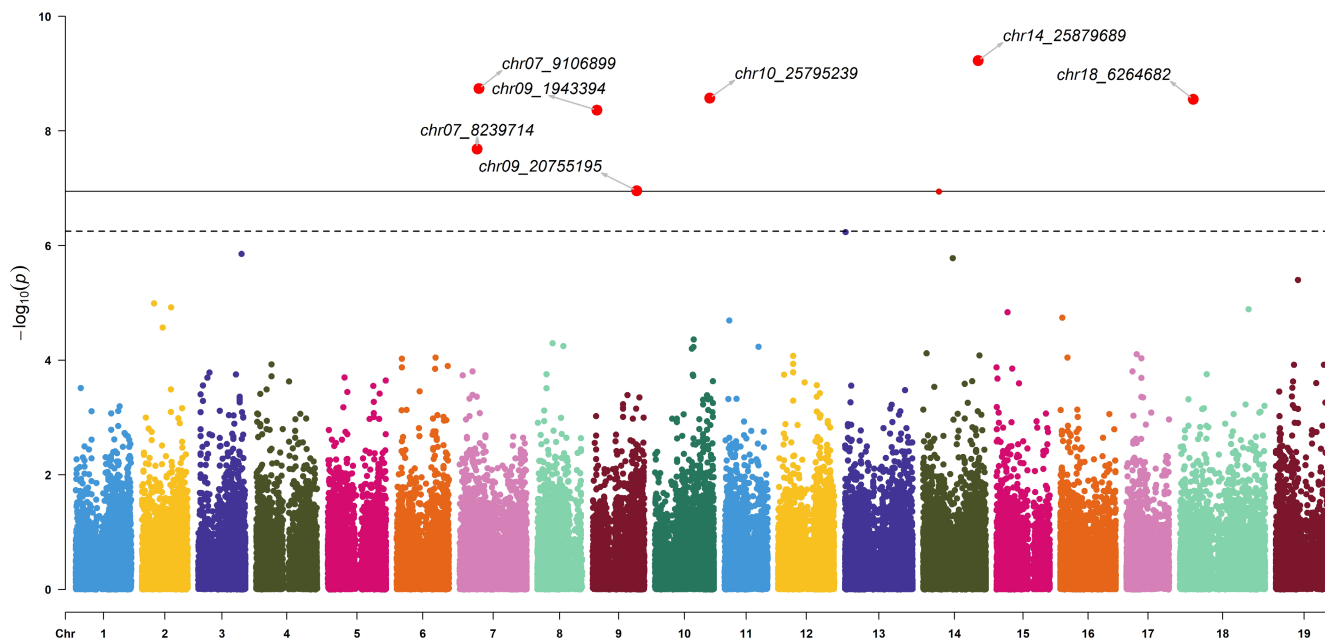


**FIGURE 7** Manhattan plot for the PC<sub>1</sub> trait (principal component from the PCA on environmental parameters). The thresholds were calculated by the Bonferroni method ( $\alpha$ /NSNP) for  $\alpha=0.05$  (dotted line) and 0.01 (solid line). Significant signals are highlighted in red. The corresponding QQ plot is presented in Figure S10a.

sampling area. The two non-admixed subpopulations could be characterized as a “northern” and a “southern” subpopulation, with a significantly different mean elevation between these two subpopulations (Figure 4).

The five clusters were well distinguished geographically (Figure S7), but elevation divided the total population into three, rather than five subpopulations (Figure S8). We, therefore,

considered a genetic structure based on two subpopulations for this geographical area (plus an admixed subpopulation). We found a significant correlation between genetic relatedness and physical distance that was explained by a phenomenon of isolation by distance, consistent with the difference in elevation between the two subpopulations. We took the significant results for IBD into account in the GEA analysis, using kinship information to control for false discovery



**FIGURE 8** Manhattan plot for the PC<sub>2</sub> trait (principal component from the PCA on environmental parameters). The thresholds were calculated by the Bonferroni method ( $\alpha$ /NSNP) for  $\alpha=0.05$  (dotted line) and 0.01 (solid line). Significant signals are highlighted in red. The corresponding QQ plot is presented in [Figure S10b](#).

rate. IBD may also have affected the STRUCTURE results, leading to the detection of false genetic subgroups (Perez et al., 2018). In this study, genetic structure was also controlled by incorporating the proportions of ancestry for each genotype from STRUCTURE.

According to Kimura's neutral theory of evolution (1968), most new mutations in a population are deleterious or neutral. The number of deleterious alleles may be higher in domesticated plants and animals than in wild species (Wallace et al., 2018). Hill–Robertson interference may result in an accumulation of deleterious mutations in genomes, reducing the efficacy of selection (Hill & Robertson, 1966). Deleterious variants are hard to predict in plants (Kono et al., 2018). We used snpEff, to predict deleterious variants and to investigate their relationship to genetic structure. We found no difference in the proportion of deleterious alleles between the *V. berlandieri* subpopulations. We were therefore unable to confirm that the admixed subpopulation had a lower genetic load due to greater mixing (Peischl et al., 2013). However, other sequence-based estimates, such as the efficacy of selection (Chen et al., 2017), may shed light on the ways in which natural selection deals with the accumulation of deleterious mutations. The link between domestication and the accumulation of deleterious mutations remains unclear, but this knowledge might be directly useful in breeding programs. Günther and Schmid (2010) classified 20% of polymorphic sites as deleterious variants in rice and *Arabidopsis* and found the genetic load to be lower in wild rice than in domesticated rice, consistent with the hypothesis of a “cost of domestication” (Lu et al., 2006). However, M.S. Kim et al. (2021) reported that the deleterious variant burden was lower in domesticated soybean than in wild soybean. Domestication has different impacts on annual and perennial plants (Gaut et al., 2015), because perennials are mostly propagated vegetatively, limiting the “cost of

domestication”. Nevertheless, vegetative propagation has resulted in the fixation of deleterious mutations in cassava (Ramu et al., 2017) and grapevine (Zhou et al., 2017), both of which are both perennials.

## 4.2 | Adaptation to the environment

The sampling area in Texas is known locally as “The Hill Country,” highlighting the alternation between valley and hills. Its vegetation changes with elevation and the various environmental pressures would also be expected to vary with elevation, accounting for the variation of a large set of environmental variables (e.g., temperature, humidity, and edaphic conditions). For this reason, despite the smaller size of the area sampled in this study than in previous studies on grapevine (Aradhyia et al., 2003; Bacilieri et al., 2013; Péros et al., 2011, 2015, 2021), we were able to highlight the existence of different climatic regions for the three subpopulations. Subpopulation 1 occurred in a hot and humid area, probably resulting in higher biotic pressure (Bois et al., 2014), with more efficient photosynthesis and secondary metabolite synthesis (Tonietto & Carbonneau, 2004). Subpopulation 2 was found in an area with a cooler climate and lower rainfall levels during the growing period, resulting in a larger water deficit. Different environmental pressures would be expected to apply in these areas of different climatic conditions, potentially leading to insular subpopulation adaptation. This result is consistent with the previous observations of Rives (1974), who reported strong phenotypic diversity between wild species, such as *V. riparia*, *V. rupestris*, and *V. berlandieri*, growing in Texas, USA, in terms of morphology, pathogen tolerance, and precocity. In particular, Rives reported that *V. berlandieri* plants from the north

TABLE 1 Markers associated with environmental parameters.

Trait	Chr	Position	n	p-Values	maf	Effect	Impact	Gene	Molecular function
Alt	chr02	16034728	281	9.60E-08	0.25	5_prime_UTR_variant	MODIFIER	Vitvi02g00558	Asparagine-tRNA ligase
Alt	chr07	13314249	281	4.61E-10	0.07	missense_variant	MODERATE	Vitvi07g02904	Multiple-splicing variant
Alt	chr15	1889550	281	4.24E-08	0.33	downstream_gene_variant	MODIFIER	Vitvi15g01070	Ferric reduction oxidase 2
PC1	chr01	16479952	277	9.20E-10	0.06	3_prime_UTR_variant	MODIFIER	Vitvi01g00842	Unknown protein
PC1	chr04	28100157	277	1.17E-09	0.08	intron_variant	MODIFIER	Vitvi04g01724	Multiple-splicing variant
PC1	chr05	21420108	277	9.64E-08	0.08	3_prime_UTR_variant	MODIFIER	Vitvi05g02116	Glycosyltransferase
PC1	chr06	503876	277	5.29E-08	0.31	missense_variant	MODERATE	Vitvi06g01533	Nitrilase
PC1	chr06	4589230	277	8.01E-08	0.47	upstream_gene_variant	MODIFIER	Vitvi06g01246	RNA binding
PC1	chr07	3341495	277	5.63E-08	0.09	downstream_gene_variant	MODIFIER	Vitvi07g01826	Metal ion binding
PC1	chr07	4179477	277	8.71E-08	0.10	downstream_gene_variant	MODIFIER	Vitvi07g01729	(BHLH domain-containing protein) Protein dimerization activity
PC1	chr19	25036767	277	9.65E-09	0.06	intron_variant	MODIFIER	Vitvi19g01666	Chromatin binding
PC2	chr07	8239714	277	2.07E-08	0.08	5_prime_UTR_variant	MODIFIER	Vitvi07g01381	Pyrimidine nucleotide-sugar Transmembrane transporter activity
PC2	chr07	9106899	277	1.81E-09	0.12	synonymous_variant	LOW	Vitvi07g02508	Nuclear localization sequence binding
PC2	chr09	1943394	277	4.32E-09	0.42	upstream_gene_variant	MODIFIER	Vitvi09g01386	Multiple-splicing variant
PC2	chr09	20755195	277	1.10E-07	0.07	missense_variant	MODERATE	Vitvi09g01590	PGG domain-containing protein
PC2	chr10	25795239	277	2.68E-09	0.19	missense_variant	MODERATE	Vitvi10g00114	Transmembrane transporter activity
PC2	chr14	25879689	277	5.90E-10	0.06	splice_region_variant&intron_variant	LOW	Vitvi14g01620	Glycosyltransferase activity
PC2	chr18	6264682	277	2.80E-09	0.09	downstream_gene_variant	MODIFIER	Vitvi18g00553	Transcription factor

had hairy leaves and ribbed shoots, whereas those from the south were almost glabrous, with smooth stems. These observations are consistent with our results revealing two gene pools in this species. In our study, the climatic parameter extracted from the Terraclimate platform followed the same directional gradient, thus, the coordinate of each genotype along the two axis of a PCA were used as indicators of the overall environmental variability. The principal components coordinates were used in a GEA and considered the environmental conditions experienced by the genotypes. The detected markers associated with environmental variability may reflect genetic regions relating to local adaptation (Williams, 1996) with potential for use as indicators for the prediction of phenotypic variation for adaptive traits (Lasky et al., 2015). The markers associated with PC1 were mostly related to frost risk during winter and spring (SFR and WFR), whereas the markers associated with PC2 were mostly associated with heat stress and water availability (HSI, HI, and DI). GEA studies

have detected associations between genetic factors and specific environmental conditions in nature, revealing markers responsible for driving local adaptation or “ecoclines” (Huxley, 1938). This method led to the detection of QTL associated with broad environmental traits in *Arabidopsis* (Frachon et al., 2018), barley (Chang et al., 2022), sorghum (Lasky et al., 2015; Menamo et al., 2021), sunflower (Todesco et al., 2020), bean (Ariani & Gepts, 2019; Elias et al., 2021), five alpine Brassicaceae species (Zulliger et al., 2013), and strawberry (Hu et al., 2022). Moreover, these markers can be genotyped in germplasm collections to improve estimates of the genetic diversity present in the collection and to select the best candidates for breeding programs. In this study, 18 SNPs associated with environmental traits were highlighted. The GEA analysis identified three QTL associated with sampling site elevation. The sequence of chr02\_16034728 is related to an asparagine-tRNA ligase (Schimmel, 1987). Asparagine is involved in nitrogen transport in plants and asparagine accumulation

is induced by multiple stresses, including mineral deficiencies, drought, salt, toxic metals, and pathogen attack (Lea et al., 2007). The chr07\_13314249 marker is associated with multiple-splicing variants and chr15\_1889550 matches with AT1G01580 in the *A. thaliana* genome, which has been shown to be related to ferric reduction oxidase 2 (Kim et al., 2019). Iron is involved in chlorophyll biosynthesis, photosynthesis, and nucleotide synthesis (Kim & Guerinot, 2007). Elevation would be expected to affect a large set of environmental variations linked to a number of different metabolic pathways. The PCA on environmental parameters extracted from the TerraClimate platform revealed associations with general environmental variability in each subpopulation. Eight QTL were identified for PC1, including chr05\_21420108, which is linked to glycosyltransferase activity (Ramasamy et al., 2005) and mostly affects sugar metabolism (Keegstra & Raikhel, 2001) and chr06\_503876, which is related to nitrilase enzyme synthesis (Piotrowski & Volmer, 2006). Nitrilases have been reported to be involved in plant-microbe interactions (Howden & Preston, 2009). chr06\_4589230 is involved in RNA binding, chr07\_3341495 is thought to be involved in metal ion binding and chr19\_25036767 is thought to be involved in chromatin binding (Gaudet et al., 2011). Seven QTL involved in diverse metabolic activities have been detected for PC2: pyrimidine nucleotide-sugar transmembrane transporter activity (chr07\_8239714, Hadley et al., 2014), nuclear localization sequence binding (chr07\_9106899, Gaudet et al., 2011), transmembrane transporter activity (chr10\_25795239, Camon et al., 2005), and a transcription factor (chr18\_6264682, Gaudet et al., 2011). Like chr05\_21420108 for PC1, chr14\_25879689 is associated with glycosyltransferase activity (Camon et al., 2005). Finally, some of the metabolic pathways identified in our GEA study, such as sugar, iron, and asparagine metabolism, may be involved in adaptation.

The genotyping of *V. berlandieri* germplasm collections for the molecular markers identified here is of potential interest for the selection of individuals carrying favorable alleles for a particular environmental stressor, to facilitate the selection of genetic resources better adapted to local cultivation environments. Here, we identified 18 QTL as correlated with environment variability; two had high  $F_{ST}$  values—0.189 and 0.221 for chr04\_28100157 (detected in GEA for PC1) and chr07\_9106899 (detected in GEA for PC2), respectively—corresponding to strong genetic differentiation between subpopulations at these loci (Weir & Clark Cockerham, 1984; Wright, 1984). However, these results require confirmation by further studies based on genetic engineering or the testing of these genotypes in a common garden.

## 5 | CONCLUSION

Our results confirm the existence of considerable genetic diversity in the *V. berlandieri* genetic background, even over a relatively small sampling area (about 40,000 km<sup>2</sup>). The analysis of genome-environment association revealed genetic markers associated with the climatic conditions in the sampling areas. These associations

should highlight genetic regions involved in the adaptation of plants to different environmental conditions. Wild genotypes constitute valuable resources that could be subjected to more precise selection and included in breeding programs. However, before their inclusion in breeding programs, these genotypes will need to undergo assessments of their agronomic performance.

## ACKNOWLEDGMENTS

We thank Benjamin Bois for providing the environmental data for each sampling site, Vera Maria Pohl for sharing her knowledge and helping with *V. berlandieri* collection and sample sequencing, Ana Clara Fanton and Julie Sappa (from Alex Edelman & Associates) for correcting the English, and Joachim Schmid, Frank Marty, and Peter Cousins for prospecting and collecting *V. berlandieri* seeds in Texas. This research was supported by the State of Hesse (Germany), Région Nouvelle Aquitaine (France), and the BAG (Bordeaux, Adelaide, and Geisenheim) alliance.

## CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest to declare.

## DATA AVAILABILITY STATEMENT

DNA sequence of the *V. berlandieri* reference genome: Data openly available in a public repository that does not issue DOIs: NCBI, BioProject ID: PRJNA886625. DNA sequences of the genotyping by sequencing of the *V. berlandieri* population: Data openly available in a public repository that does not issue DOIs: NCBI, BioProject ID: PRJNA886619. Other data: Data available on request from the authors.

## ORCID

Louis Blois  <https://orcid.org/0000-0001-8392-5639>

Marina de Miguel  <https://orcid.org/0000-0001-6398-2660>

## REFERENCES

- Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, 5(1), 170191. <https://doi.org/10.1038/sdata.2017.191>
- Abebe, T. D., Naz, A. A., & Léon, J. (2015). Landscape genomics reveal signatures of local adaptation in barley (*Hordeum vulgare* L.). *Frontiers in Plant Science*, 6, 813. <https://doi.org/10.3389/fpls.2015.00813>
- Aguirre-Liguori, J. A., Morales-Cruz, A., & Gaut, B. S. (2022). Evaluating the persistence and utility of five wild *Vitis* species in the context of climate change. *Molecular Ecology*, 31(24), 6457–6472. <https://doi.org/10.1111/mec.16715>
- Aradhya, M. K., Dangl, G. S., Prins, B. H., Jean-Michel Boursiquot, M., Walker, A., Meredith, C. P., & Simon, C. J. (2003). Genetic structure and differentiation in cultivated grape, *Vitis vinifera* L. *Genetics Research*, 81(3), 179–192. <https://doi.org/10.1017/s0016672303006177>
- Ariani, A., & Gepts, P. (2019). Signatures of Environmental Adaptation During Range Expansion of Wild Common Bean (*Phaseolus vulgaris*). *March*. <https://doi.org/10.1101/571042>
- Arroyo-García, R., Ruiz-García, L., Bolling, L., Ocete, R., López, M. A., Arnold, C., Ergul, A., Söylemezoğlu, G., Uzun, H. I., Cabello, F.,

- Ibáñez, J., Aradhya, M. K., Atanassov, A., Atanassov, I., Balint, S., Cenis, J. L., Costantini, L., Goris-Lavets, S., Grando, M. S., ... Martinez-Zapater, J. M. (2006). Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *sativa*) based on chloroplast DNA polymorphisms. *Molecular Ecology*, 15(12), 3707–3714. <https://doi.org/10.1111/j.1365-294X.2006.03049.x>
- Bacilieri, R., Lacombe, T., Le Cunff, L., Di Vecchi-Staraz, M., Laucou, V., Genna, B., Péros, J.-P., This, P., & Boursiquot, J.-M. (2013). Genetic structure in cultivated grapevines is linked to geography and human selection. *BMC Plant Biology*, 13(1), 25. <https://doi.org/10.1186/1471-2229-13-25>
- Bianchi, D., Caramanico, L., Grossi, D., Brancadoro, L., & De Lorenzis, G. (2020). How do novel M-rootstock (*Vitis* spp.) genotypes cope with drought? *Plants (Basel)*, 9(10), 1385. <https://doi.org/10.3390/plant9101385>
- Bois, B., Moriondo, M., & Jones, G. V. (2014). Thermal Risk Assessment for Viticulture Using Monthly Temperature Data. In *Xth International Terroir Congress Tokaj Eger Hungary* (pp. 227–32).
- Bois, B., Zito, S., & Calonne, A. (2017). Climate vs grapevine pests and diseases worldwide: The first results of a global survey. *OENO One*, 51(2), 133–139. <https://doi.org/10.20870/oeno-one.2017.51.2.1780>
- Boubals, D. (1966). Etude de La Distribution et Des Causes de La Résistance Au Phylloxera Radicicole Chez Les Vitacées, no. 16: 145–84.
- Bragg, J. G., Supple, M. A., Andrew, R. L., & Borevitz, J. O. (2015). Genomic variation across landscapes: Insights and applications. *The New Phytologist*, 207(4), 953–967. <https://doi.org/10.1111/nph.13410>
- Branas, J., Bernon, G., & Levadoux, L. (1946). *Éléments de viticulture générale*. École National d'Agriculture de Montpellier.
- Buiteveld, J., Vendramin, G. G., Leonardi, S., Kamer, K., & Geburek, T. (2007). Genetic diversity and differentiation in European beech (*Fagus sylvatica* L.) stands varying in management history. *Forest Ecology and Management*, 247(1–3), 98–106. <https://doi.org/10.1016/j.foreco.2007.04.018>
- Camon, E. B., Barrell, D. G., Dimmer, E. C., Lee, V., Magrane, M., Maslen, J., Binns, D., & Apweiler, R. (2005). An evaluation of GO annotation retrieval for BioCreAtIvE and GOA. *BMC Bioinformatics*, 6(Suppl 1), S17. <https://doi.org/10.1186/1471-2105-6-S1-S17>
- Canaguier, A., Grimplet, J., Di Gaspero, G., Scalabrin, S., Duchêne, E., Choisne, N., Mohellibi, N., Guichard, C., Rombauts, S., Le Clainche, I., Bérard, A., Chauveau, A., Bounon, R., Rustenholz, C., Morgante, M., Le Paslier, M. C., Brunel, D., & Adam-Blondon, A. F. (2017). A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.V3). *Genomics Data*, 14(December), 56–62. <https://doi.org/10.1016/j.gdata.2017.09.002>
- Carbonneau, A. (1998). Aspects qualitatifs, 258–276. In J. R. Tiercelin (Ed.), *Traité d'irrigation* (p. 1011). Tec et Doc Lavoisier.
- Chang, C.-W., Fridman, E., Mascher, M., Himmelbach, A., & Schmid, K. (2022). Physical geography, isolation by distance and environmental variables shape genomic variation of wild barley (*Hordeum vulgare* L. ssp. *spontaneum*) in the southern levant. *Heredity*, 128(2), 107–119. <https://doi.org/10.1038/s41437-021-00494-x>
- Chen, J., Glémin, S., & Lascoux, M. (2017). Genetic diversity and the efficacy of purifying selection across plant and animal species. *Molecular Biology and Evolution*, 34(6), 1417–1428. <https://doi.org/10.1093/molbev/msx088>
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Xiangyi, L., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of drosophila melanogaster strain W1118; Iso-2; Iso-3. *Fly*, 6(2), 80–92. <https://doi.org/10.4161/fly.19695>
- Cipriani, G., Spadotto, A., Jurman, I., Di Gaspero, G., Crespan, M., Meneghetti, S., Frare, E., Vignani, R., Cresti, M., Morgante, M., Pezzotti, M., Pe, E., Policriti, A., & Testolin, R. (2010). The SSR-based molecular profile of 1005 grapevine (*Vitis vinifera* L.) accessions uncovers new synonymy and parentages, and reveals a large admixture amongst varieties of different geographic origin. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, 121(8), 1569–1585. <https://doi.org/10.1007/s00122-010-1411-9>
- Condon, A. G., Richards, R. A., Rebetzke, G. J., & Farquhar, G. D. (2004). Breeding for high water-use efficiency. *Journal of Experimental Botany*, 55(407), 2447–2460. <https://doi.org/10.1093/jxb/erh277>
- Cormier, F., Lawac, F., Maledon, E., Gravillon, M.-C., Nudol, E., Mournet, P., Vignes, H., Chair, H., & Arnau, G. (2019). A reference high-density genetic map of greater yam (*Dioscorea alata* L.). *Theoretical and Applied Genetics*, 132(6), 1733–1744. <https://doi.org/10.1007/s00122-019-03311-6>
- Cortés, A. J., & López-Hernández, F. (2021). Harnessing crop wild diversity for climate change adaptation. *Genes*, 12(5), 783. <https://doi.org/10.3390/genes12050783>
- Cortés, A. J., López-Hernández, F., & Blair, M. W. (2022). Genome-environment associations, an innovative tool for studying heritable evolutionary adaptation in orphan crops and wild relatives. *Frontiers in Genetics*, 13(August), 910386. <https://doi.org/10.3389/fgene.2022.910386>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- de Andrés, M. T., Cabezas, J. A., Cervera, M. T., Borrego, J., Martínez-Zapater, J. M., & Jouve, N. (2007). Molecular characterization of grapevine rootstocks maintained in germplasm collections. *American Journal of Enology and Viticulture*, 58(1), 75–86.
- De La Torre, A. R., Wilhite, B., & Neale, D. B. (2019). Environmental genome-wide association reveals climate adaptation is shaped by subtle to moderate allele frequency shifts in loblolly pine. *Genome Biology and Evolution*, 11(10), 2976–2989. <https://doi.org/10.1093/gbe/evz220>
- Di Vecchi-Staraz, M., Laucou, V., Bruno, G., Lacombe, T., Gerber, S., Bourse, T., Boselli, M., & This, P. (2009). Low level of pollen-mediated gene flow from cultivated to wild grapevine: Consequences for the evolution of the endangered subspecies *Vitis vinifera* L. subsp. *silvestris*. *Journal of Heredity*, 100(1), 66–75. <https://doi.org/10.1093/jhered/esn084>
- Ekhvaia, J., Gurushidze, M., Blattner, F. R., & Akhalkatsi, M. (2014). Genetic diversity of *Vitis vinifera* in Georgia: Relationships between local cultivars and wild grapevine, *V. vinifera* L. subsp. *silvestris*. *Genetic Resources and Crop Evolution*, 61(8), 1507–1521. <https://doi.org/10.1007/s10722-014-0125-2>
- Elias, J. C. F., Gonçalves-Vidigal, M. C., Ariani, A., Valentini, G., da Conceição Martiniano-Souza, M., Bisneta, M. V., & Gepts, P. (2021). Genome-environment association analysis for bio-climatic variables in common bean (*Phaseolus vulgaris* L.) from Brazil. *Plants*, 10(8), 1572. <https://doi.org/10.3390/plants10081572>
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, 6(5), e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Ergül, A., Perez-Rivera, G., Söylemezoğlu, G., Kazan, K., & Arroyo-García, R. (2011). Genetic diversity in Anatolian wild grapes (*Vitis vinifera* subsp. *silvestris*) estimated by SSR markers. *Plant Genetic Resources*, 9(3), 375–383. <https://doi.org/10.1017/S1479262111000013>
- Évanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14(8), 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Flutre, T., Bacilieri, R., Bécavin, I., Berger, G., Bertrand, Y., Boursiquot, J.-M., Fodor, A., Lacombe, T., Laucou, V., Launay, A., Le Cunff, L.,

- Romieu, C., This, P., Péros, J.-P., & Doligez, A. (2019). Genome-wide association study of a diverse grapevine panel: Example of berry weight. *Acta Horticulturae*, 1248, 227–234.
- Frachon, L., Bartoli, C., Carrère, S., Bouchez, O., Chaubet, A., Gautier, M., Roby, D., & Roux, F. (2018). A genomic map of climate adaptation in *Arabidopsis thaliana* at a micro-geographic scale. *Frontiers in Plant Science*, 9, 967. <https://doi.org/10.3389/fpls.2018.00967>
- Francia, E., Tacconi, G., Crosatti, C., Barabaschi, D., Bulgarelli, D., Dall'Aglio, E., & Valè, G. (2005). Marker assisted selection in crop plants. *Plant Cell, Tissue and Organ Culture*, 82(3), 317–342. <https://doi.org/10.1007/s11240-005-2387-z>
- Frenkel, O., Portillo, I., Brewer, M. T., Péros, J. P., Cadle-Davidson, L., & Milgrom, M. G. (2012). Development of microsatellite markers from the transcriptome of *Erysiphe necator* for analysing population structure in North America and Europe. *Plant Pathology*, 61(1), 106–119. <https://doi.org/10.1111/j.1365-3059.2011.02502.x>
- Galet, P. (1956). *Dictionnaire encyclopédique des cépages et de leurs synonymes*. Libre & Solidaire.
- Galet, P. (1988). *Cépages et Vignobles de France – Tome 1 Les Vignes Américaines*. IMPRIMERIE Charles DEHAN.
- Gaudet, P., Livstone, M. S., Lewis, S. E., & Thomas, P. D. (2011). Phylogenetic-based propagation of functional annotations within the gene ontology consortium. *Briefings in Bioinformatics*, 12(5), 449–462. <https://doi.org/10.1093/bib/bbr042>
- Gaut, B. S., Díez, C. M., & Morrell, P. L. (2015). Genomics and the contrasting dynamics of annual and perennial domestication. *Trends in Genetics*, 31(12), 709–719. <https://doi.org/10.1016/j.tig.2015.10.002>
- Girollet, N., Rubio, B., Lopez-Roques, C., Valière, S., Ollat, N., & Bert, P.-F. (2019). *De novo* phased assembly of the *Vitis riparia* grape genome. *Scientific Data*, 6, 127. <https://doi.org/10.1038/s41597-019-0133-3>
- González-Martínez, S. C., Ridout, K., & Pannell, J. R. (2017). Range expansion compromises adaptive evolution in an outcrossing plant. *Current Biology*, 27(16), 2544–2551.e4. <https://doi.org/10.1016/j.cub.2017.07.007>
- Grassi, F., Imazio, S., Failla, O., Scienza, A., Ocete Rubio, R., Lopez, M. A., Sala, F., & Labra, M. (2003). Genetic isolation and diffusion of wild grapevine Italian and Spanish populations as estimated by nuclear and chloroplast SSR analysis. *Plant Biology*, 5(6), 608–614. <https://doi.org/10.1055/s-2003-44689>
- Günther, T., & Schmid, K. J. (2010). Deleterious amino acid polymorphisms in *Arabidopsis thaliana* and rice. *Theoretical and Applied Genetics*, 121(1), 157–168. <https://doi.org/10.1007/s00122-010-1299-4>
- Hadley, B., Maggioni, A., Ashikov, A., Day, C. J., Haselhorst, T., & Tiralongo, J. (2014). Structure and function of nucleotide sugar transporters: Current progress. *Computational and Structural Biotechnology Journal*, 10(16), 23–32. <https://doi.org/10.1016/j.csbj.2014.05.003>
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: AK-means clustering algorithm. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, 28(1), 100–108. <https://doi.org/10.2307/2346830>
- Hill, W. G., & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetics Research*, 8(3), 269–294. <https://doi.org/10.1017/S0016672300010156>
- Hill, W. G., & Weir, B. S. (1988). Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical Population Biology*, 33(1), 54–78. [https://doi.org/10.1016/0040-5809\(88\)90004-4](https://doi.org/10.1016/0040-5809(88)90004-4)
- Howden, A. J. M., & Preston, G. M. (2009). Nitrilase enzymes and their role in plant-microbe interactions. *Microbial Biotechnology*, 2(4), 441–451. <https://doi.org/10.1111/j.1751-7915.2009.00111.x>
- Hu, Y., Feng, C., Yang, L., Edger, P. P., & Kang, M. (2022). Genomic population structure and local adaptation of the wild strawberry *Fragaria nilgerrensis*. *Horticulture Research*, 9(January), uhab059. <https://doi.org/10.1093/hr/uhab059>
- Huang, M., Liu, X., Zhou, Y., Summers, R. M., & Zhang, Z. (2018). BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience*, 8(2), giy154. <https://doi.org/10.1093/gigascience/giy154>
- Huxley, J. (1938). Clines: An auxiliary taxonomic principle. *Nature*, 142(3587), 219–220. <https://doi.org/10.1038/142219a0>
- Imazio, S., Labra, M., Grassi, F., Scienza, A., & Failla, O. (2006). Chloroplast microsatellites to investigate the origin of grapevine. *Genetic Resources and Crop Evolution*, 53(5), 1003–1011. <https://doi.org/10.1007/s10722-004-6896-0>
- Jones, G. V. (2006). Climate and Terroir: Impacts of Climate Variability and Change on Wine. <https://www.semanticscholar.org/paper/Climate-and-Terroir-%3A-Impacts-of-Climate-and-Change-Jones/8062ae1bbc0d8338ff3249d846c5de58f7eb0f39>
- Kang, H. M., Sul, J. H., Susan K. Service, Zaitlen, N. A., Kong, S.-y., Freimer, N. B., Sabatti, C., & Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, 42(4), 348–354. <https://doi.org/10.1038/ng.548>
- Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., & Eskin, E. (2008). Efficient control of population structure in model organism association mapping. *Genetics*, 178(3), 1709–1723. <https://doi.org/10.1534/genetics.107.080101>
- Keegstra, K., & Raikhel, N. (2001). Plant glycosyltransferases. *Current Opinion in Plant Biology*, 4(3), 219–224. [https://doi.org/10.1016/S1369-5266\(00\)00164-3](https://doi.org/10.1016/S1369-5266(00)00164-3)
- Kim, M.-S., Lozano, R., Kim, J. H., Bae, D. N., Kim, S.-T., Park, J.-H., Choi, M. S., Kim, J., Ok, H.-C., Park, S.-K., Gore, M. A., Moon, J.-K., & Jeong, S.-C. (2021). The patterns of deleterious mutations during the domestication of soybean. *Nature Communications*, 12(1), 97. <https://doi.org/10.1038/s41467-020-20337-3>
- Kim, S. A., & Guerinot, M. L. (2007). Mining iron: Iron uptake and transport in plants. *FEBS Letters*, 581(12), 2273–2280. <https://doi.org/10.1016/j.febslet.2007.04.043>
- Kim, S. A., LaCroix, I. S., Gerber, S. A., & Guerinot, M. L. (2019). The iron deficiency response in *Arabidopsis thaliana* requires the phosphorylated transcription factor URI. *Proceedings of the National Academy of Sciences*, 116(50), 24933–24942. <https://doi.org/10.1073/pnas.1916892116>
- Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, 217(5129), 624–626. <https://doi.org/10.1038/217624a0>
- Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511623486>
- Kono, T. J. Y., Lei, L., Shih, C.-H., Hoffman, P. J., Morrell, P. L., & Fay, J. C. (2018). Comparative genomics approaches accurately predict deleterious variants in plants. *G3: Genes, Genomes, Genetics*, 8(10), 3321–3329. <https://doi.org/10.1534/g3.118.200563>
- Lasky, J. R., Upadhyaya, H. D., Ramu, P., Santosh Deshpande, C., Hash, T., Bonnette, J., Juenger, T. E., Hyma, K., Acharya, C., Mitchell, S. E., Buckler, E. S., Brenton, Z., Kresovich, S., & Morris, G. P. (2015). Genome-environment associations in sorghum landraces predict adaptive traits. *Science Advances*, 1(6), e1400218. <https://doi.org/10.1126/sciadv.1400218>
- Laucou, V., Lacombe, T., Dechesne, F., Siret, R., Bruno, J.-P., Dessup, M., Dessup, T., Ortigosa, P., Parra, P., Roux, C., Santoni, S., Varès, D., Péros, J. P., Boursiquot, J. M., & This, P. (2011). High throughput analysis of grape genetic diversity as a tool for germplasm collection management. *Theoretical and Applied Genetics*, 122(6), 1233–1245. <https://doi.org/10.1007/s00122-010-1527-y>
- Laucou, V., Launay, A., Bacilieri, R., Lacombe, T., Adam-Blondon, A.-F., Bérard, A., Chauveau, A., de Andrés, M. T., Hausmann, L., Ibáñez, J., le Paslier, M.-C., Maghradze, D., Martínez-Zapater, J. M., Maul, E., Ponnaiah, M., Töpfer, R., Péros, J.-P., & Boursiquot, J.-M. (2018). Extended diversity analysis of cultivated grapevine *Vitis vinifera* with 10K genome-wide SNPs. *PLoS One*, 13(2), e0192540. <https://doi.org/10.1371/journal.pone.0192540>
- Lea, P. J., Sodek, L., Parry, M. A. J., Shewry, P. R., & Halford, N. G. (2007). Asparagine in plants. *Annals of Applied Biology*, 150(1), 1–26. <https://doi.org/10.1111/j.1744-7348.2006.00104.x>



- Liang, Z., Duan, S., Sheng, J., Zhu, S., Ni, X., Shao, J., Liu, C., Nick, P., du, F., Fan, P., Mao, R., Zhu, Y., Deng, W., Yang, M., Huang, H., Liu, Y., Ding, Y., Liu, X., Jiang, J., ... Dong, Y. (2019). Whole-genome resequencing of 472 *Vitis* accessions for grapevine diversity and demographic history analyses. *Nature Communications*, 10(1), 1190. <https://doi.org/10.1038/s41467-019-09135-8>
- Lopes, M. S., Mendonça, D., Rodrigues dos Santos, M., Eiras-Dias, J. E., & da Câmara Machado, A. (2009). New insights on the genetic basis of Portuguese grapevine and on grapevine domestication. *Genome*, 52(9), 790–800. <https://doi.org/10.1139/g09-048>
- Lu, J., Tang, T., Tang, H., Huang, J., Shi, S., & Chung-I, W. (2006). The accumulation of deleterious mutations in rice genomes: A hypothesis on the cost of domestication. *Trends in Genetics*, 22(3), 126–131. <https://doi.org/10.1016/j.tig.2006.01.004>
- Marrano, A., Micheletti, D., Lorenzi, S., Neale, D., & Stella Grando, M. (2018). Genomic signatures of different adaptations to environmental stimuli between wild and cultivated *Vitis vinifera* L. *Horticulture Research*, 5, 34. <https://doi.org/10.1038/s41438-018-0041-2>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, 17(1), 10–12. <https://doi.org/10.14806/ej.17.1.200>
- Martinez, L. E., Cavagnaro, P. F., Masuelli, R. W., & Zúñiga, M. (2006). SSR-based assessment of genetic diversity in south American *Vitis vinifera* varieties. *Plant Science*, 170(6), 1036–1044. <https://doi.org/10.1016/j.plantsci.2005.12.006>
- McGovern, P. (2003). *Ancient Wine: The Search for the Origins of Viniculture*. *Ancient Wine: The Search for the Origins of Viniculture*. <https://doi.org/10.1515/9781400849536.1>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytisky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Menamo, T., Kassahun, B., Borrell, A. K., Jordan, D. R., Tao, Y., Hunt, C., & Mace, E. (2021). Genetic diversity of Ethiopian sorghum reveals signatures of climatic adaptation. *Theoretical and Applied Genetics*, 134(2), 731–742. <https://doi.org/10.1007/s00122-020-03727-5>
- Morales-Cruz, A., Aguirre-Liguori, J. A., Zhou, Y., Minio, A., Riaz, S., Walker, A. M., Cantu, D., & Gaut, B. S. (2021). Introgression among north American wild grapes (*Vitis*) fuels biotic and abiotic adaptation. *Genome Biology*, 22(1), 254. <https://doi.org/10.1186/s13059-021-02467-z>
- Myles, S., Boyko, A. R., Owens, C. L., Brown, P. J., Grassi, F., Aradhya, M. K., Prins, B., Reynolds, A., Chia, J.-M., Ware, D., Bustamante, C. D., & Buckler, E. S. (2011). Genetic structure and domestication history of the grape. *Proceedings of the National Academy of Sciences*, 108(9), 3530–3535. <https://doi.org/10.1073/pnas.1009363108>
- Myles, S., Chia, J.-M., Hurwitz, B., Simon, C., Zhong, G. Y., Buckler, E., & Ware, D. (2010). Rapid genomic characterization of the genus *Vitis*. *PLoS One*, 5(1), e8219. <https://doi.org/10.1371/journal.pone.0008219>
- Nordborg, M., & Tavaré, S. (2002). Linkage disequilibrium: What history has to tell us. *Trends in Genetics*, 18(2), 83–90. [https://doi.org/10.1016/s0168-9525\(02\)02557-x](https://doi.org/10.1016/s0168-9525(02)02557-x)
- Ohta, T. (1973). Slightly deleterious mutant substitutions in evolution. *Nature*, 246(5428), 96–98. <https://doi.org/10.1038/246096a0>
- Parmesan, C., & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421(6918), 37–42. <https://doi.org/10.1038/nature01286>
- Peischl, S., Dupanloup, I., Kirkpatrick, M., & Excoffier, L. (2013). On the accumulation of deleterious mutations during range expansions. *Molecular Ecology*, 22(24), 5972–5982. <https://doi.org/10.1111/mec.12524>
- Perez, M. F., Franco, F. F., Bombonato, J. R., Bonatelli, I. A. S., Khan, G., Romeiro-Brito, M., Fegies, A. C., Ribeiro, P. M., Silva, G. A. R., & Moraes, E. M. (2018). Assessing population structure in the face of isolation by distance: Are we neglecting the problem? *Diversity and Distributions*, 24(12), 1883–1889. <https://doi.org/10.1111/ddi.12816>
- Péros, J.-P., Berger, G., Portemont, A., Boursiquot, J.-M., & Lacombe, T. (2011). Genetic variation and biogeography of the disjunct *Vitis* subg. *Vitis* (Vitaceae). *Journal of Biogeography*, 38(3), 471–486. <https://doi.org/10.1111/j.1365-2699.2010.02410.x>
- Péros, J.-P., Cousins, P., Launay, A., Cubry, P., Walker, A., Prado, E., Peressotti, E., Wiedemann-Merdinoglu, S., Laucou, V., Merdinoglu, D., This, P., Boursiquot, J.-M., & Doligez, A. (2021). Genetic diversity and population structure in *Vitis* species illustrate phylogeographic patterns in eastern North America. *Molecular Ecology*, 30(10), 2333–2348. <https://doi.org/10.1111/mec.15881>
- Péros, J.-P., Launay, A., Berger, G., Lacombe, T., & This, P. (2015). MybA1 gene diversity across the *Vitis* genus. *Genetica*, 143(3), 373–384. <https://doi.org/10.1007/s10709-015-9836-3>
- Piotrowski, M., & Volmer, J. J. (2006). Cyanide metabolism in higher plants: Cyanoalanine hydratase is a NIT4 homolog. *Plant Molecular Biology*, 61(1–2), 111–122. <https://doi.org/10.1007/s11103-005-6217-9>
- Pluess, A. R., Frank, A., Heiri, C., Lalagüe, H., Vendramin, G. G., & Oddou-Muratorio, S. (2016). Genome–environment association study suggests local adaptation to climate at the regional scale in *Fagus sylvatica*. *New Phytologist*, 210(2), 589–601. <https://doi.org/10.1111/nph.13809>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959.
- Ramasamy, V., Ramakrishnan, B., Boeggeman, E., Ratner, D. M., Seeberger, P. H., & Qasba, P. K. (2005). Oligosaccharide preferences of Beta1,4-galactosyltransferase-I: Crystal structures of Met340His mutant of human Beta1,4-galactosyltransferase-I with a Pentasaccharide and Trisaccharides of the N-glycan moiety. *Journal of Molecular Biology*, 353(1), 53–67. <https://doi.org/10.1016/j.jmb.2005.07.050>
- Ramu, P., Esuma, W., Kawuki, R., Rabbi, I. Y., Egesi, C., Bredeson, J. V., Bart, R. S., Verma, J., Buckler, E. S., & Fei, L. (2017). Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. *Nature Genetics*, 49(6), 959–963. <https://doi.org/10.1038/ng.3845>
- Ravaz, L. (1902). *LES VIGNES AMERICAINES: porte greffes et producteurs directs, caracteres, aptitudes*. S.I.
- Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, 24(17), 4348–4370. <https://doi.org/10.1111/mec.13322>
- Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., Kresovich, S., Goodman, M. M., & Buckler, E. S. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences*, 98(20), 11479–11484. <https://doi.org/10.1073/pnas.201394398>
- Riaz, S., Pap, D., Uretsky, J., Laucou, V., Boursiquot, J.-M., Kocsis, L., & Andrew Walker, M. (2019). Genetic diversity and parentage analysis of grape rootstocks. *Theoretical and Applied Genetics*, 132(6), 1847–1860. <https://doi.org/10.1007/s00122-019-03320-5>
- Richards, C. M., Volk, G. M., Reilley, A. A., Henk, A. D., Lockwood, D. R., Reeves, P. A., & Forsline, P. L. (2009). Genetic diversity and population structure in *malus sieversii*, a wild progenitor species of domesticated apple. *Tree Genetics & Genomes*, 5(2), 339–347. <https://doi.org/10.1007/s11295-008-0190-9>
- Riou, C., Pieri, P., & Le Clech, B. (1994). Consommation d'eau de la vigne en conditions hydriques non limitantes. Formulation simplifiée de la transpiration. *Vitis*, 33, 109–115.

- Rives, M. (1974). Les Vignes Sauvages Comme Sources de Gènes Pour l'amélioration. *VITIS – Journal of Grapevine Research*, 13(3), 186. <https://doi.org/10.5073/vitis.1974.13.186-197>
- Root, T. L., Price, J. T., Hall, K. R., Schneider, S. H., Rosenzweig, C., & Alan Pounds, J. (2003). Fingerprints of global warming on wild animals and plants. *Nature*, 421(6918), 57–60. <https://doi.org/10.1038/nature01333>
- Santure, A. W., & Garant, D. (2018). Wild GWAS—Association mapping in natural populations. *Molecular Ecology Resources*, 18(4), 729–738. <https://doi.org/10.1111/1755-0998.12901>
- Schatz, M. C., Witkowski, J., & Richard McCombie, W. (2012). Current challenges in de novo plant genome sequencing and assembly. *Genome Biology*, 13(4), 243. <https://doi.org/10.1186/gb4015>
- Schimmel, P. (1987). Aminoacyl TRNA synthetases: General scheme of structure-function relationships in the polypeptides and recognition of transfer RNAs. *Annual Review of Biochemistry*, 56(1), 125–158. <https://doi.org/10.1146/annurev.bi.56.070187.001013>
- Schmid, J., Manty, F., & Cousins, P. (2009). Collecting *Vitis berlandieri* from native habitat sites. *Acta Horticulturae*, 827(May), 151–154. <https://doi.org/10.17660/ActaHortic.2009.827.22>
- Tello, J., & Ibáñez, J. (2023). Review: Status and prospects of association mapping in grapevine. *Plant Science*, 327, 111539. <https://doi.org/10.1016/j.plantsci.2022.111539>
- The UniProt Consortium. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49(D1), D480–D489. <https://doi.org/10.1093/nar/gkaa1100>
- This, P., Lacombe, T., & Thomas, M. R. (2006). Historical origins and genetic diversity of wine grapes. *Trends in Genetics*, 22(9), 511–519. <https://doi.org/10.1016/j.tig.2006.07.008>
- Todesco, M., Owens, G. L., Bercovich, N., Légaré, J.-S., Soudi, S., Burge, D. O., Huang, K., Ostevik, K. L., Drummond, E. B. M., Imerovski, I., Lande, K., Pascual-Robles, M. A., Nanavati, M., Jahani, M., Cheung, W., Staton, S. E., Muñoz, S., Nielsen, R., Donovan, L. A., ... Rieseberg, L. H. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*, 584(7822), 602–607. <https://doi.org/10.1038/s41586-020-2467-6>
- Tonietto, J., & Carbonneau, A. (2004). A multicriteria climatic classification system for grape-growing regions worldwide. *Agricultural and Forest Meteorology*, 124(1), 81–97. <https://doi.org/10.1016/j.agrfor.2003.06.001>
- Tuberosa, R., Salvi, S., Giuliani, S., Sanguineti, M. C., Bellotti, M., Conti, S., & Landi, P. (2007). Genome-wide approaches to investigate and improve maize response to drought. *Crop Science*, 47(S3), S-120–S-141. <https://doi.org/10.2135/cropsci2007.04.0001IPBS>
- Vadez, V., Kholova, J., Medina, S., Kakkera, A., & Anderberg, H. (2014). Transpiration efficiency: New insights into an old story. *Journal of Experimental Botany*, 65(21), 6141–6153. <https://doi.org/10.1093/jxb/eru040>
- Voss-Fels, K., Frisch, M., Qian, L., Kontowski, S., Friedt, W., Gottwald, S., & Snowdon, R. J. (2015). Subgenomic diversity patterns caused by directional selection in bread wheat gene pools. *The Plant Genome*, 8(2). <https://doi.org/10.3835/plantgenome2015.03.0013>
- Wallace, J. G., Rodgers-Melnick, E., & Buckler, E. S. (2018). On the road to breeding 4.0: Unraveling the good, the bad, and the boring of crop quantitative genomics. *Annual Review of Genetics*, 52(1), 421–444. <https://doi.org/10.1146/annurev-genet-120116-024846>
- Wang, Z., Kang, M., Liu, H., Gao, J., Zhang, Z., Li, Y., Rongling, W., & Pang, X. (2014). High-level genetic diversity and complex population structure of Siberian apricot (*Prunus sibirica* L.) in China as revealed by nuclear SSR markers. *PLoS One*, 9(2), e87381. <https://doi.org/10.1371/journal.pone.0087381>
- Weir, B. S., & Clark Cockerham, C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6), 1358–1370. <https://doi.org/10.2307/2408641>
- Williams, G. C. (1996). *Adaptation and natural selection: A critique of some current evolutionary thought*. Facsim. ed., with New pref. Princeton Science Library. Princeton University Press.
- Wright, S. (1984). *Evolution and the genetics of populations, volume 2: Theory of gene frequencies*. University of Chicago Press. <https://press.uchicago.edu/ucp/books/book/chicago/E/bo5961634.html>
- Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., Bradbury, P. J., Yu, J., Arnett, D. K., Ordovas, J. M., & Buckler, E. S. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, 42(4), 355–360. <https://doi.org/10.1038/ng.546>
- Zhou, Y., Massonnet, M., Sanjak, J. S., Cantu, D., & Gaut, B. S. (2017). Evolutionary genomics of grape (*Vitis vinifera* ssp. *vinifera*) domestication. *Proceedings of the National Academy of Sciences of the United States of America*, 114(44), 11715–11720. <https://doi.org/10.1073/pnas.1709257114>
- Zulliger, D., Schnyder, E., & Gugerli, F. (2013). Are adaptive loci transferable across genomes of related species? Outlier and environmental association analyses in alpine Brassicaceae species. *Molecular Ecology*, 22(6), 1626–1639. <https://doi.org/10.1111/mec.12199>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Blois, L., de Miguel, M., Bert, P.-F., Girollet, N., Ollat, N., Rubio, B., Segura, V., Voss-Fels, K. P., Schmid, J., & Marguerit, E. (2023). Genetic structure and first genome-wide insights into the adaptation of a wild relative of grapevine, *Vitis berlandieri*. *Evolutionary Applications*, 16, 1184–1200. <https://doi.org/10.1111/eva.13566>