Horticulture, Environment, and Biotechnology (2025) 66:1719–1740
https://doi.org/10.1007/s13580-025-00749-0

RESEARCH REPORT

Exploring genetic diversity in North American *Vitis* species: screening for resistance loci and implications for utilization in molecular breeding programs

Geovani Luciano de Oliveira[1] · Paola Bettinelli[2] · Daniela Nicolini[2] · Anete Pereira de Souza[1,3] · Mara Fernandes Moura Furlan[4] · Marco Stefanini[2] · Silvia Vezzulli[2]

Received: 26 March 2025 / Revised: 23 July 2025 / Accepted: 26 July 2025 / Published online: 4 September 2025

Abstract

North American *Vitis* species serve as a vital reservoir of genetic variation, offering valuable resources for molecular breeding programs focused on developing cultivars with enhanced resistance, adaptability, and quality traits for sustainable viticulture. The effective conservation and utilization of collections involving these species require a thorough understanding of their genetic diversity, population structure, and gene flow. In this study, 323 North American *Vitis* accessions categorized into six species groups were genotyped using 29 polymorphic microsatellite (SSR) markers. The objectives were to assess the genetic diversity and population structure, as well as to establish a representative core collection. Additionally, SSR markers associated with 15 resistance loci (*R*-loci) were analyzed to identify potential resistance to downy mildew, powdery mildew, black rot, Pierce's disease, and phylloxera. The analysis revealed high genetic diversity, with 643 alleles identified, an expected heterozygosity (H_E) of 0.86, and an observed heterozygosity (H_O) of 0.74. Nine genetic groups were identified, with clear evidence of a substructure within some species. A core collection comprising 95 accessions capable of retaining all SSR alleles detected in the entire collection was established. Characterization of the *R*-loci revealed that 123 wild genotypes carried a single *R*-locus, 36 carried two *R*-loci, and 10 harbored three *R*-loci associated with pathogen resistance. The richness evident in the studied genetic pool represents an extensive reservoir of underexplored genetic diversity and crossbreeding potential. These findings have the potential to bolster the sustainable management, conservation, and subsequent molecular breeding applications of wild *Vitis* resources amidst emerging challenges in viticulture.

Keywords Grapevine · SSR markers · Genetic resources · Resistance donors · Core collection · Population structure

Communicated by Jin Hoe Huh

✉ Silvia Vezzulli
silvia.vezzulli@fmach.it

Geovani Luciano de Oliveira
geovani@unicamp.br

Paola Bettinelli
paola.bettinelli@fmach.it

Daniela Nicolini
daniela.nicolini@fmach.it

Anete Pereira de Souza
anete@unicamp.br

Mara Fernandes Moura Furlan
mara.moura@sp.gov.br

Marco Stefanini
marco.stefanini@fmach.it

[1] Molecular Biology and Genetic Engineering Center (CBMEG), Universidade Estadual de Campinas (UNICAMP), Av. Cândido Rondon, n 400, Campinas, SP, Brazil

[2] Grapevine Physiology and Breeding Unit, Research and Innovation Centre, Fondazione Edmund Mach, Via Edmund Mach 1, 38098 San Michele all'Adige, TN, Italy

[3] Department of Plant Biology, Biology Institute (IB), Universidade Estadual de Campinas (UNICAMP), R. Monteiro Lobato, n 255, Campinas, SP, Brazil

[4] Advanced Fruit Research Center, Agronomic Institute (IAC), Luiz Pereira dos Santos Avenue, n 1.500, Jundiaí, SP, Brazil

1 Introduction

The domesticated grapevine (*Vitis vinifera* L.) is the most widely cultivated grape species; however, its productivity has historically been constrained by its susceptibility to pests and diseases (Dry et al. 2019). Additionally, factors such as the restricted genetic diversity of commercially relevant *V. vinifera* cultivars and the widespread reliance on a small number of rootstocks have resulted in increased interest in wild *Vitis* species as valuable resources for viticulture (Migicovsky et al. 2016; Aguirre-Liguori et al. 2022). Most wild *Vitis* species exhibit substantial genetic diversity, which represents a valuable reservoir of alleles that could significantly contribute to the improvement of global viticulture (Atak 2024). Through coevolution with a wide range of pathogens, North American wild grapevines are believed to have developed diverse resistance strategies. As a result, they have been continuously used in crossbreeding programs as parental lines to increase natural disease resistance in both scion cultivars and rootstocks (Vezzulli et al. 2022; Péros et al. 2023). In recent years, the pressures related to climate change have highlighted the urgent need to identify additional germplasms for rootstock and scion breeding, especially due to the fact that grapevines are perennial crops that propagate clonally, which limits their ability to undergo rapid evolution (Heinitz et al. 2019; Aguirre-Liguori et al. 2022).

The potential value of different *Vitis* species must first be determined to establish the viability of their use in crossbreeding programs. The prebreeding process involves all activities related to identifying desirable traits and genes in unadapted germplasms, such as wild species, and transferring them into well-adapted genetic backgrounds (Sharma et al. 2013). One of the main challenges in practical prebreeding involves the determination of which activities are necessary and how many activities are required to evaluate exotic material for identifying subsets of donors that are likely to harbor novel and useful genetic variation for breeding (Sukumaran et al. 2022). In practice, the phenotypic screening of large germplasm collections in field or greenhouse environments for all traits of interest is unfeasible due to logistical and financial limitations. The development of representative subsets of reduced sizes is a widely used strategy when germplasm collections are too large for practical use. The objective of these core collections is to identify allelic diversity and optimize characterization and evaluation efforts by focusing on a subset of accessions (Boccacci et al. 2021).

DNA markers, such as simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs), are now widely used for developing core collections, analyzing genetic diversity, determining genetic relationships, and conducting DNA fingerprinting, even in the early stages of plant development (Žulj Mihaljević et al. 2020; de Oliveira et al. 2020; Wang et al. 2021; Kaya et al. 2023; Tao et al. 2023). These markers accurately represent the genetic diversity of the entire collection and help circumvent issues related to incomplete provenance information and environmental interactions, as polymorphisms are directly detected at the DNA level and are free from environmental influences (Boccacci et al. 2021; Amiteye 2021). The identification of genetic markers associated with agronomic and fruit quality traits enables the use of marker-assisted selection (MAS) during prebreeding, thereby allowing for screens of large plant populations and progenies. This approach provides important supplementary information, thus facilitating the identification of individuals carrying beneficial alleles for strategic crossbreeding combinations (Collard and Mackill 2008; De Lorenzis et al. 2022).

In grapevines, the screening of resistance loci (*R*-loci) containing genes associated with pathogen resistance is especially effective for the targeted selection of parental lines with the potential to combine multiple *R*-loci in their offspring, thus increasing both the level and durability of resistance (Eibach and Töpfer 2015; Zini et al. 2019; Vezzulli et al. 2019; Possamai et al. 2020). Downy mildew (DM) and powdery mildew (PM) have been the most extensively studied biotic stress factors in quantitative trait loci (QTL) mapping, thereby leading to a vast and continually expanding list of *R*-loci. In contrast, for other types of biotic stress factors, such as phylloxera, Pierce's disease (PD), and black rot (BR), the number of identified *R*-loci remains limited. To date, 37 loci associated with DM resistance (*Rpv*) and 18 loci associated with PM resistance (*Ren/Run*) have been reported. For phylloxera resistance (*Rdv*), eight loci have been described, with only three loci being linked to BR resistance (*Rgb*), and one locus being linked to PD resistance (*Pdr*) (VIVC 2024).

Typically, population structure studies represent the first step in examining sample sets, thus providing a foundation for subsequent genetic analyses or allowing for the inference of the origins of individuals with unknown population characteristics, especially when population admixture is possible (Porras-Hurtado et al. 2013). When considering that North American *Vitis* species of the subgenus *Euvitis* are dioecious and interfertile, species delimitation becomes confusing, which is partly due to the lack of consensus among systematic botanists regarding what defines a true species and partly due to the extreme morphological variation observed within species (Wan et al. 2013; Heinitz et al. 2019; Péros et al. 2023). Molecular methods, which complement traditional morphological approaches, have been increasingly employed to address these taxonomic

challenges and aid in *Vitis* systematics (Myles et al. 2010; Ma et al. 2018; Péros et al. 2023).

Considering the importance of wild *Vitis* species for developing new pathogen-resistant cultivars and identifying super-donor parents of resistance, the Fondazione Edmund Mach (FEM) maintains a germplasm bank with accessions from different North American *Vitis* species with the potential for incorporation into molecular breeding programs. However, information on the origin of this material is limited, and the species classification of the accessions is based on the female parent plant during seed collection. Without information on the number of seeds originating from the same cluster or female parent plant, establishing the genetic relationships between accessions and assessing the diversity preserved in the collection becomes challenging. Thus, the objective of this study was to describe the level of diversity and genetic structure of North American *Vitis* species in this germplasm bank using SSR markers. Additionally, we aimed to determine the composition of a core collection representative of the entire allelic variation. Finally, a secondary objective was to conduct *R*-loci screening to identify potential resistance donors for strategic crossbreeding combinations. Overall, the results obtained represent an essential prebreeding step, guiding the effective use of genetic diversity in this collection and lay the groundwork for the incorporation of wild grapevine species into molecular breeding programs.

2 Materials and methods

2.1 Plant material and DNA extraction

A total of 323 grapevine accessions from the Fondazione Edmund Mach (FEM) wild germplasm collection in San Michele all'Adige, Italy (46°11'N, 11°08'E, 223 m asl) were utilized in this study (Supplementary Table S1). These accessions were subsequently integrated into the historical collection described by Emanuelli et al. (2013) and originated from seeds collected from female vines in the University of California, Davis (USA) germplasm collection. The seeds from open-pollinated female parents were categorized into six North American species: 50 *Vitis aestivalis*, 61 *Vitis arizonica*, 90 *Vitis cinerea*, 98 *Vitis monticola*, 12 *Vitis mustangensis*, 8 *Vitis rotundifolia*, and 4 unknown species (Fig. 1). The germplasm sub-collection was established in 2018 with a spacing of 2.5 m (interrow) × 1 m (intrarow), and the vines were trained using the vertical shoot positioning (VSP) method. Due to the mortality of some accessions, the age of the vines differed depending on the introduction and regeneration dates of the original plants. They were maintained according to local standard commercial practices for weed and pest control, fertilization, and pruning.

Genomic DNA was extracted from young leaf material homogenized in a TissueLyser (Qiagen, Valencia, CA, USA) using the cetyltrimethylammonium bromide (CTAB) protocol (Doyle 1991) in 96-well plates, with minor modifications. The buffer consisted of 2% CTAB, 1400 mM NaCl, 2% 2-mercaptoethanol, 20 mM EDTA, 200 mM Tris-HCl (pH 8.0), and 2% w/v polyvinylpyrrolidone (PVP-40). The solution was preheated, and the samples were incubated at 60 °C in a water bath for 1 h, followed by two washes with

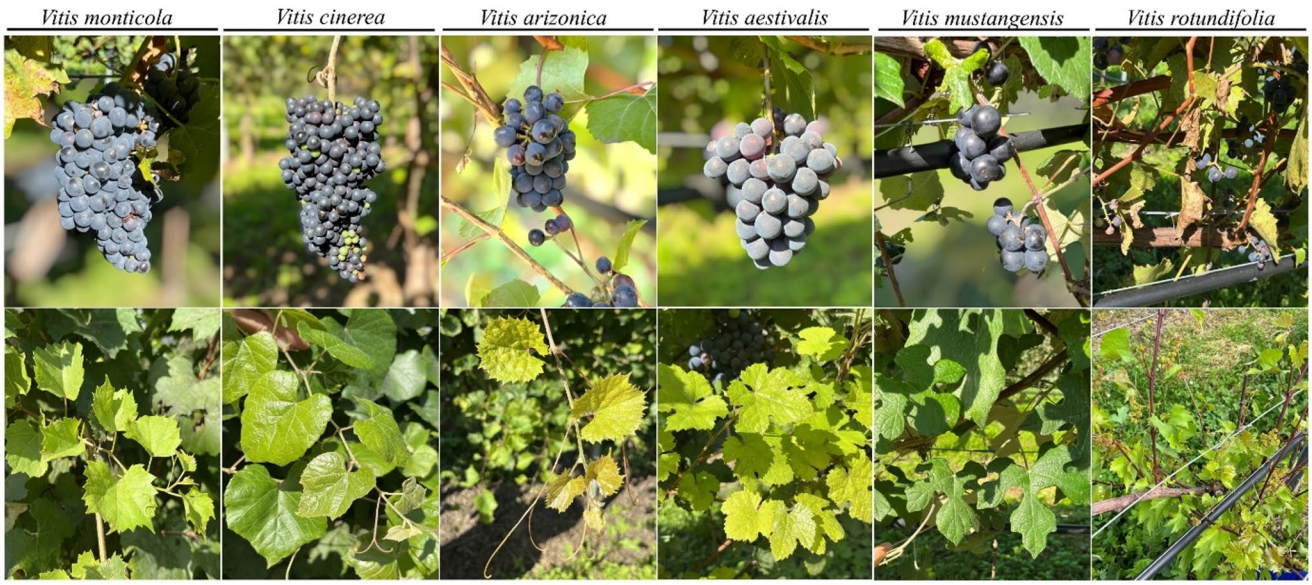

Fig. 1 Representative accessions of the utilized *Vitis* species groups in this study

chloroform–isoamyl alcohol (24:1). A 2/3 volume of cold isopropanol and 10 µL of 3 M NaAc per sample were then added. The samples were washed twice with 70% ethanol, and finally, the pellet was dried and resuspended in 100 µL of ultrapure water. The DNA quality was checked via 1% agarose gel electrophoresis, the DNA concentration was quantified using a NanoDrop 8000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), and the DNA concentration was normalized to 10 ng/µL.

2.2 Microsatellite analysis

A set of 29 SSR markers was used to genotype the wild accessions, including the following nine reference SSRs selected by the international scientific community for universal grapevine identification according to the GenRes081 and GrapeGen06 EU projects (This et al. 2004; Maul et al. 2012): VVS2 (Thomas and Scott 1993), VVMD5, VVMD7 (Bowers et al. 1996), VVMD25, VVMD27, VVMD28, VVMD32 (Bowers et al. 1999), VrZAG62, and VrZAG79 (Sefc et al. 1999). The remaining 20 SSRs were identified from an analysis of *R*-loci based on their distribution across the genome and locus quality: UDV025, UDV111, UDV095 (Di Gaspero et al. 2005), GF13-9, GF13-1 (Zhang et al. 2009), VVIP26 (Merdinoglu et al. 2005), GF05-13 (Fechter et al. 2014), UDV360 (Venuti et al. 2013), UDV737, UDV305 (Di Gaspero et al. 2012), GF09-46, GF09-47 (Schwander et al. 2012), Sc47-18, Sc08_0071_014 (Coleman et al. 2009), CenGen6 (van Heerden et al. 2014), ScorGF15-32 (Zendler et al. 2017), GF14-42 (Rex et al. 2014), VMC2C3 (Goto-Yamamoto et al. 2006), SC34-8 (NCBI: GF111545), and VVCh14-77 (Riaz et al. 2018b). Additional information on the loci selected for genetic analysis (including SSR primer sequences) is available in Supplementary Table S2.

The forward primers of all the loci were labeled with distinct fluorescent dyes (6-FAM, HEX, or NED). The *R*-loci markers were amplified using the KAPA2G Fast Multiplex PCR Kit 2X (Kapa Biosystems, Düren, Germany) in 10 µL reaction mixtures containing 5 µL of master mix, 100 pmol of each primer, and 1 ng of template DNA. Amplifications of the nine reference loci were performed in a 10 µL final volume using the Qiagen Multiplex Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. All PCRs were conducted using a GeneAmp 9700 thermal cycler, with the protocols provided in Supplementary Table S2. The 'Pinot Noir' and 'Teroldego' varieties were used as reference genotypes because of their well-characterized allelic profiles for the analyzed loci. Capillary electrophoresis of all SSR loci was performed using an ABI 3130xl Genetic Analyzer (Life Technologies, Foster City, CA, USA). The fragments (alleles) were sized using GeneMarker software v3.0 (SoftGenetics, State College, PA, USA), with the GeneScan 500 LIZ size standard used as an internal ladder (Life Technologies, Foster City, CA, USA). The complete SSR genotyping dataset, including allele sizes for all of the loci and accessions, is provided in Supplementary Table S1.

2.3 Genetic diversity and population structure analyses

Genetic diversity was estimated both within species groups and across the entire germplasm. Descriptive statistics based on the genotyping data were generated by using GenAlEx 6.5 (Peakall and Smouse 2012) to assess the number of alleles per locus (Na), effective number of alleles (Ne), observed heterozygosity (H_O), expected heterozygosity (H_E), Shannon's information index (I), fixation index (F), and private alleles (Pa). Additionally, the null allele frequency (*r*) and polymorphism information content (PIC) were determined by using CERVUS 3.0.7 (Kalinowski et al. 2007). Discrimination power (D*j*) values were also assessed to compare the effectiveness of the SSRs in identifying and differentiating varieties (Tessier et al. 1999). An identity analysis was conducted by using CERVUS, with a minimum matching criterion of 20 loci used and an allowance for one fuzzy match, in order to prevent a duplicate analysis of identical genotypes.

A Bayesian model-based cluster analysis was performed by using STRUCTURE 2.3.4 software (Pritchard et al. 2000) to estimate the approximate number of clusters (K) within the dataset and to assign individuals to the most suitable cluster. The STRUCTURE settings were configured to disregard population information and to apply an admixture model with correlated allele frequencies. Additionally, a "hierarchical structure analysis" (Vähä et al. 2007) was conducted, and STRUCTURE was separately utilized in a second round of analysis for clusters that exhibited potential substructures, as indicated by other genetic structure analyses.

All of the simulations employed the admixture model, with 100,000 replicates used as burn-in and 1,000,000 replicates used for the Markov chain Monte Carlo (MCMC) processes. The number of clusters (K) ranged from 1 to 15 in the first round and from 1 to 10 in the second round, with 10 replicate runs being conducted to assess the variation in likelihood for each K. The most likely K value was selected by using the Evanno method (Evanno et al. 2005), as implemented in the Structure Harvester software (Earl and vonHoldt 2012). The best alignment across the 10 runs for the optimal K values was achieved by using the greedy algorithm in CLUMPP 1.1.2 (Jakobsson and Rosenberg 2007), and the results were visualized with DISTRUCT software 1.1 (Rosenberg 2004). The genotypes were classified into

genetic groups based on a membership probability threshold (q) of ≥0.70.

Discriminant analysis of principal components (DAPC) was conducted by using the R package *adegenet*, which applies a nonparametric approach that does not rely on the Hardy–Weinberg equilibrium assumption (Jombart et al. 2010). This analysis was employed to illustrate the genetic divergence among clusters, enabling the identification of related clusters. The *find.clusters* function was employed to identify the number of clusters within the wild germplasm by performing successive rounds of K-means clustering with an increasing number of clusters (K), with 20 set as the maximum. The optimal cluster number was determined based on the Bayesian information criterion (BIC). The DAPC outcomes are displayed as multidimensional scaling plots.

Genetic relationships among the accessions were evaluated by using a distance-based model estimated via Rogers' genetic distance (Rogers 1972). The resulting distance matrix was utilized to construct a dendrogram employing the neighbor-joining (NJ) algorithm (Saitou and Nei 1987) with 1,000 bootstrap replicates implemented in the R package *poppr* (Kamvar et al. 2014). The final dendrogram was generated by using iTOL 6.8.2 (Letunic and Bork 2021).

2.4 Construction of a core collection

The R package *Core Hunter v.3* (De Beukelaer et al. 2018) was used to generate a core collection that maximizes the representation of alleles that are present in the full collection within a reduced number of accessions. Different samples were created by adjusting the size parameter of the core collections to identify subsets of genotypes capable of capturing 100% of allele coverage. The sampling fractions ranged from 0.1 to 0.3 for the entire collection. The genetic diversity parameters of the resulting core collections were then compared to those of the full dataset by using GenAlEx v.6.5 (Peakall and Smouse 2012).

2.5 R-loci analysis

Using the "all-vs-all" approach (Zini et al. 2019; Vezzulli et al. 2019), all of the accessions were examined for 15 screenable and reliable *R*-loci: six loci linked to DM resistance (*Rpv1*, *Rpv3*, *Rpv10*, *Rpv12*, *Rpv14*, and *Rpv27*), five loci associated with PM resistance (*Run1*, *Ren1*, *Ren2*, *Ren3*, and *Ren9*), two loci correlated with BR resistance (*Rgb1* and *Rgb3*), one locus associated with PD resistance (*PdR1*), and one locus linked to Phylloxera resistance (*Rdv1*). The resistance gene symbols, causal agent names for the diseases (traits), markers related to resistance, associated resistance alleles, and genotypes of origin are detailed in Table 1. In this study, the absence of *R*-loci specifically refers only to *R*-loci with exploitable associated markers, without excluding the possibility of known loci lacking reliable markers or novel/unknown loci.

3 Results

3.1 Genetic diversity

No redundancies were detected among the 323 *Vitis* accessions analyzed by using 29 SSR markers, which yielded 323 distinct molecular profiles (Supplementary Table S1). The mean Na was 22.24, ranging from 6 alleles (GF09-47) to 38 alleles (UDV305), with a total of 643 alleles across the entire collection (Table 2). The Ne varied from 1.55 (GF09-47) to 16.48 (UDV305), with a mean value of 8.97.

Across the entire germplasm, the analysis revealed a high mean H_E value, also referred to as the genetic diversity index (0.86), ranging from 0.36 (GF09-47) to 0.94 (UDV305). The mean H_O was 0.74, ranging from 0.31 (GF09-47) to 0.93 (GF13-9). H_E was higher than H_O at 24 loci and very similar at the remaining 5 loci. Among these 24 loci ($He > H_O$), the probability of null alleles (*r*) was significantly high (>0.20) for four loci (VVMD32, GF05-13, CenGen6 and SC34-8). However, the exclusion of these loci did not substantially affect the inference of genetic structure when compared to the complete marker set (Supplementary Fig. S3). The PIC estimates ranged from 0.32 (GF09-47) to 0.94 (UDV305), with a mean value of 0.84. The D*j* value was greater than 0.80 for 27 of the 29 loci, with the highest value calculated for the UDV305 locus (0.95). Overall, 24 loci showed high values for both the PIC and D*j* (>0.80). Among the 643 SSR alleles identified, 28.4% had frequencies greater than 5% and were categorized as common alleles; 35.6% had frequencies between 1 and 5% and were classified as less common alleles; and 36.0% had frequencies below 1% and were considered rare alleles.

Statistical indices were calculated for each subset to assess the genetic diversity within each species group (Table 3). The average Na ranged from 3.79 in the unknown species group to 13.55 in *V. monticola*, whereas the average Ne varied from 2.39 in *V. mustangensis* to 5.60 in *V. cinerea* var. *helleri*. The *V. mustangensis* group also presented the lowest values for H_E (0.49) and Shannon's information index (I, 0.91), whereas the highest values for H_E and I were observed for *V. rotundifolia* (0.84) and *V. aestivalis* var. *aestivalis* (1.86), respectively. The fixation index (F) ranged from −0.11 for *V. cinerea* var. *floridana* to 0.21 for *V. rotundifolia*.

Among the 643 observed alleles, 505 were shared among the species groups, whereas the remaining 138 represented

Table 1 Resistance (*R*) loci with causal agents, chromosomes, associated markers, corresponding resistance alleles/haplotypes, and reference genotypes. Additional information about the *R*-loci is available in the vitis international variety catalog (VIVC) (www.vivc.de, session: data on breeding genetics) and Bettinelli et al. (2023a)

Symbol	Resistance Trait	Chromosome	Associated Marker	Resistant Allele/Haplotype	Reference Genotype
Rpv1-Run1	*Plasmopara viticola/Erysiphe necator*	12	VMC4f3.1	186	VRH3082-1-42
			VMC8g9	160	
			Sc34-8	216	
			Sc35-2	238	
Rpv3-1	*Plasmopara viticola*	18	UDV305	299	Regent
			UDV737	279	
Rpv3-2			UDV305	null	Souvignier gris
			UDV737	297	
Rpv3-3			UDV305	null	Merzling
			UDV737	271	
$Rpv3^{321-312}$			UDV305	321	Chancellor
			UDV737	312	
$Rpv3^{361-299}$			UDV305	361	Eger 28
			UDV737	299	
$Rpv3^{299-314}$			UDV305	299	Courderc 13
			UDV737	314	
$Rpv3^{null-287}$			UDV305	null	Chancellor
			UDV737	287	
Rpv10		9	GF09-44	230	Severnyi
			GF09-46	416	
			GF09-47	299	
Rpv12		14	UDV340	197	Kunleany
			UDV345	236	
			UDV360	227	
Rpv14		5	GF05-13	294	Börner
			UDV111	114	
Rpv27		18	VVCS1H077H16R1-1	96	Norton
			UDV737	294	
Ren1	*Erysiphe necator*	13	Sc47-18	249	Kishmish vatkana
			SC8_0071_014	143	
Ren2		14	VVIP26	144	NY95.03
			UDV025	125	
Ren3		15	ScORGF15-02	242	Regent
Ren9		15	CenGen6	287	Regent
Rgb1	*Phyllosticta ampelicida*	14	GF14-42	332	Merzling
			UDV095	123	
Rgb3		14	VVIP22	335	Merzling
			VMC2C3	159	
Rdv1	*Daktulosphaira vitifoliae*	13	GF13-01	213	Börner
			GF13-09	339	
PdR1	*Xylella fastidiosa*	14	VVCh14-77	202/204	*V. arizonica*
			VVIP26	146–146	

private alleles (Pa) specific to distinct species. The *V. monticola* and *V. cinerea* var. *helleri* subsets exhibited the greatest number of private alleles, with 36 and 26, respectively. In contrast, the *V. mustangensis* group and the unknown species group demonstrated the lowest numbers of Pa at 8 and 5, respectively.

3.2 Genetic structure and relationships

STRUCTURE analysis was conducted to determine the probable number of genetic groups (K) within the dataset, and ancestral values per genotype for K values ranging from 1 to 15 were recorded. Using the Evanno method (Evanno et al. 2005), the most likely number of populations was determined to be K = 8, as it presented the highest

Table 2 Genetic parameters of the 29 microsatellite loci analyzed across 323 *Vitis* spp. Accessions, including the number of alleles per locus (Na), effective number of alleles (Ne), observed heterozygosity (H_O), expected heterozygosity (H_E), polymorphism information content (PIC), discrimination power (D*j*), and estimated null allele frequency (*r*)

Locus	Na	Ne	H_O	H_E	PIC	D*j*	*r*
VVS2	22	8.97	0.84	0.89	0.88	0.89	0.03
VVMD5	26	13.02	0.73	0.92	0.92	0.93	0.12
VVMD7	18	7.92	0.64	0.87	0.86	0.88	0.16
VVMD25	21	11.07	0.92	0.91	0.90	0.91	-0.01
VVMD27	24	11.46	0.92	0.91	0.91	0.92	-0.01
VVMD28	30	12.68	0.71	0.92	0.92	0.92	0.13
VVMD32	23	6.21	0.54	0.84	0.83	0.84	0.23
VRZAG62	28	9.91	0.84	0.90	0.89	0.90	0.04
VRZAG79	19	7.49	0.80	0.87	0.85	0.87	0.04
UDV025	22	7.12	0.87	0.86	0.85	0.86	-0.01
VVIP26	36	14.75	0.90	0.93	0.93	0.94	0.01
VVCh14-77	26	12.06	0.88	0.92	0.91	0.92	0.02
UDV111	20	8.48	0.84	0.88	0.87	0.88	0.03
GF05-13	28	12.37	0.58	0.92	0.91	0.92	0.22
UDV360	16	5.02	0.81	0.80	0.78	0.80	-0.01
UDV737	36	13.37	0.86	0.93	0.92	0.93	0.03
UDV305	38	16.48	0.70	0.94	0.94	0.95	0.15
GF13-9	25	13.40	0.93	0.93	0.92	0.93	-0.01
GF13-1	16	7.54	0.85	0.87	0.86	0.87	0.01
Sc47-18	20	6.15	0.82	0.84	0.82	0.84	0.01
Sc08_0071_014	24	7.77	0.83	0.87	0.86	0.87	0.02
CenGen6	29	10.53	0.52	0.91	0.90	0.92	0.28
ScorGF15-32	8	5.14	0.66	0.81	0.78	0.81	0.10
SC34-8	8	3.14	0.42	0.68	0.64	0.68	0.24
GF09-46	20	6.12	0.66	0.84	0.82	0.84	0.12
GF09-47	6	1.55	0.31	0.36	0.32	0.36	0.10
GF14-42	20	6.44	0.71	0.84	0.83	0.85	0.09
UDV095	19	8.94	0.78	0.89	0.88	0.89	0.06
VMC2C3	15	5.05	0.72	0.80	0.78	0.80	0.05
Total	643	260.15					
Mean	22.17	8.97	0.74	0.86	0.84	0.86	
SE[a]	1.46	0.67	0.03	0.02	0.02	0.02	

[a]Standard errors of the mean values

Table 3 Descriptive statistics of the mean loci values observed for each *Vitis* species group, including the sample size (N), number of alleles (Na), shannon's information index (I), observed heterozygosity (H_O), expected genetic diversity (H_E), fixation index (F), and private alleles (Pa)

Species Groups	*N*	Na	Ne	I	H_O	H_E	F	Pa
Vitis aestivalis	24	9.90	4.29	1.67	0.73	0.73	-0.02	11
Vitis aestivalis var. *aestivalis*	26	11.34	5.01	1.86	0.79	0.78	-0.02	13
Vitis arizonica	61	11.38	3.88	1.61	0.75	0.71	-0.08	18
Vitis cinerea var. *floridana*	22	8.93	3.68	1.51	0.73	0.66	-0.11	11
Vitis cinerea var. *helleri*	68	11.90	5.60	1.82	0.76	0.76	0.01	26
Vitis monticola	98	13.55	4.39	1.76	0.77	0.75	-0.05	36
Vitis mustangensis	12	4.03	2.39	0.91	0.44	0.49	0.03	8
Vitis rotundifolia	8	7.66	5.59	1.80	0.62	0.84	0.21	10
unknown	4	3.79	3.24	1.20	0.70	0.77	-0.06	5
Mean	35.89	9.16	4.23	1.57	0.70	0.72	-0.01	15.33
SE[a]	1.86	0.28	0.12	0.03	0.02	0.01	0.02	3.10

[a]Standard errors of the mean values

ΔK value (Fig. 2). After establishing the threshold for the group assignment at q≥0.70, 273 accessions (84.5%) were unequivocally assigned to a cluster at K=8; the remaining 50 accessions (15.5%) exhibited membership values lower than 0.70 and were considered admixed genotypes. At this level of structure (K=8), a clear separation was observed among the sampled accessions based on species, with an evident substructure within certain species (Fig. 3a).

The *V. monticola* accessions were distributed between the groups VMO1 (*n*=36) and VMO2 (*n*=34), with 70 accessions showing high association values (q≥0.70) and 28 accessions exhibiting intermediate associations (q<0.70), thus classified as admixed. The VAR group was associated with the *V. arizonica* species, with 49 accessions showing q≥0.70. The VMU group (*n*=12) consisted exclusively of *V. mustangensis* accessions, whereas the VAE group (*n*=29) was associated with *V. aestivalis* var. *aestivalis*. The VA-VC group was unique in exhibiting a predominance of multiple species, comprising accessions of *V. aestivalis* (21), *V. cinerea* var. *floridana* (21), and *V. cinerea* var. *helleri* (10) simultaneously. The remaining *V. cinerea* var. *helleri* accessions were distributed between the VCH1 (*n*=29) and VCH2 (*n*=30) groups. The eight accessions classified as *V. rotundifolia* were not clustered together: three were classified as admixed, whereas the others were distributed among the VMO1, VAE, and VCH2 groups. Among the accessions of unknown classification, two were identified as admixed and two were assigned to the VAE group.

A second round of the STRUCTURE analysis was conducted for the VA-VC group due to its heterogeneous composition (Fig. 3b), with the aim of identifying potential substructures, as suggested by the other clustering analyses (Figs. 4 and 5). The highest ΔK value was observed at K=2 (Supplementary Fig. S1), indicating the presence of two subgroups within this group. Essentially, the *V. cinerea* var. *floridana* accessions formed the VCF subgroup (21), all of which showed high membership values (q≥0.70), whereas the VA-VC2 subgroup (31) consisted of the remaining *V. aestivalis* and *V. cinerea* var. *helleri* accessions from the VA-VC group.

Furthermore, DAPC was conducted with no prior information about the grouping of the evaluated accessions. An analysis of the BIC values indicated that the partitioning of the accessions into nine clusters was the most plausible approach for explaining the variance within this set of genotypes (Supplementary Fig. S2). In the preliminary data transformation step, DAPC was able to explain 99% of the total genetic variation through the maintenance of 150 principal components (PCs). The DAPC scatter plot, which is based on the first and second discriminant functions, depicts the distributions of the nine groups (Fig. 4), thus revealing significant genetic differentiation among most of the groups and low variance within the groups. The color

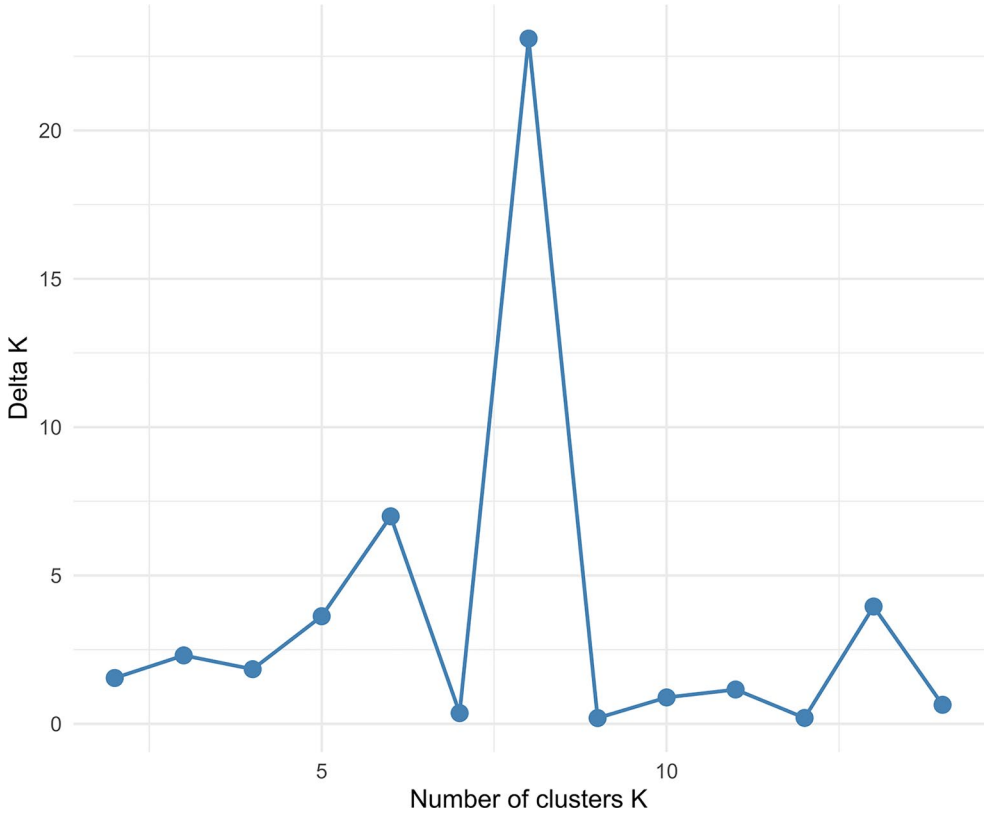

Fig. 2 The most likely number of genetic clusters (K) within the full dataset of 323 accessions was determined by using the method described by Evanno et al. (2005). K values from ten separate runs, ranging from 1 to 15, are shown. The delta K graph shows the maximum value at K=8

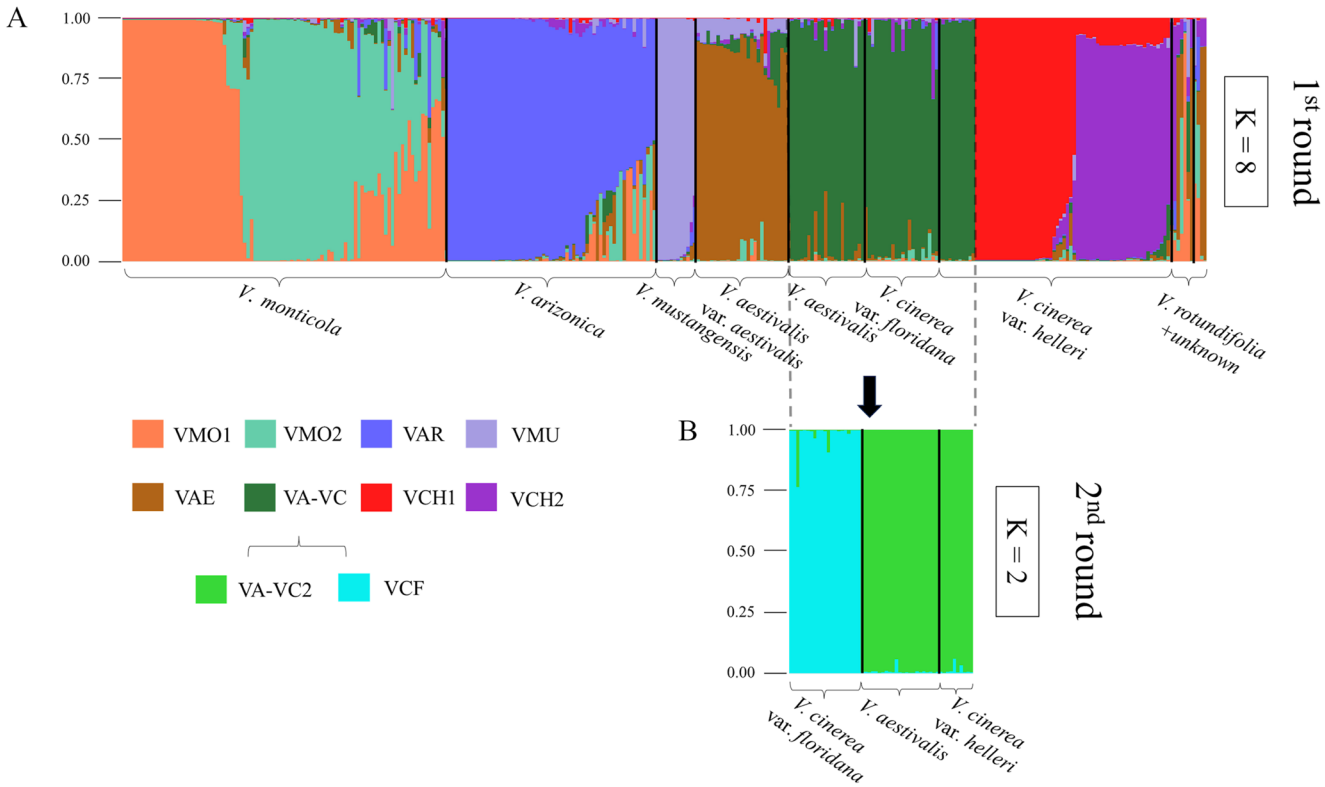

Fig. 3 Bar plot of the estimated membership coefficients (q) for the 323 analyzed *Vitis* accessions. Each vertical bar represents an individual, divided into colored segments that indicate the proportional membership in each inferred cluster. Groups were labeled based on the predominant species within the accessions. (**A**) First round of the STRUCTURE analysis, showing the inferred genetic structure for K = 8. Clusters VMO1 and VMO2: *V. monticola*; Cluster VAR: *V. arizonica;* Cluster VMU: *V. mustangensis*; Cluster VAE: *V. aestivalis* var. *aestivalis*; Cluster VA-VC: *V. aestivalis*, *V. cinerea* var *floridana*, and *V. cinerea* var. *helleri*; Clusters VCH1 and VCH2: *V. cinerea* var. *helleri*. (**B**). Second round of the STRUCTURE analysis showing substructuring within the VA-VC cluster for K = 2. Cluster VCF: *V. cinerea* var *floridana*; Cluster VA-VC2: *V. aestivalis* and *V. cinerea* var. *helleri*

coding from the STRUCTURE assignment was used in the DAPC visualization. The allocation of individuals into clusters based on the DAPC results showed strong similarity to those obtained from the Bayesian clustering analysis. Both analyses revealed the same clustering pattern, reflecting the gene pools of the species. Specifically, DAPC clusters 1 through 9 corresponded to the STRUCTURE groups VMO2, VA-VC2, VCF, VAR, VMU, VMO1, VAE, VCH1, and VCH2, respectively. Similar to what was observed in the STRUCTURE analysis, the *V. rotundifolia* accessions did not form a single group in the DAPC and were distributed across clusters 1, 6, 7, and 9. Moreover, the accessions with unknown species were allocated to clusters 1 and 7.

Finally, a neighbor-joining (NJ) clustering analysis was performed to further explore the genetic relationships among the grapevine genotypes (Fig. 5). Overall, the resulting dendrogram exhibited a pattern consistent with the previous analyses, providing a clear visualization of clustering based on the species group classification. The dendrogram highlighted *V. monticola* as being genetically more distant from the other species, with its accessions forming two distinct major clusters. Within the large branch comprising the remaining species, *V. arizonica* emerged as the most genetically distinct. Additionally, above the species level, closer genetic relationships were observed between *V. mustangensis* and *V. aestivalis* var. *aestivalis*, as well as among *V. cinerea* var. *helleri*, *V. cinerea* var. *floridana*, and *V. aestivalis*. The genetic similarity analysis indicated that the accession Arizonica Carlin-125 (ID: 125), previously classified as an unidentified species, clustered with accessions identified as *V. monticola*, suggesting a potential taxonomic assignment. The NJ dendrogram also revealed some cases where the initial taxonomic classifications were inconsistent with their placement in the dendrogram. A total of nine accessions were flagged as potentially misclassified within the collection. Among these, three accessions classified as *V. rotundifolia* were distributed into different clusters: two (IDs 123 and 124) grouped with *V. monticola* accessions, and one (ID 274) clustered with *V. aestivalis* var. *aestivalis*. Furthermore, two *Vitis aestivalis* accessions (IDs 238 and 246) grouped within the *V. aestivalis* var. *aestivalis* cluster, and two *V. cinerea* var. *helleri* accessions (IDs 315 and

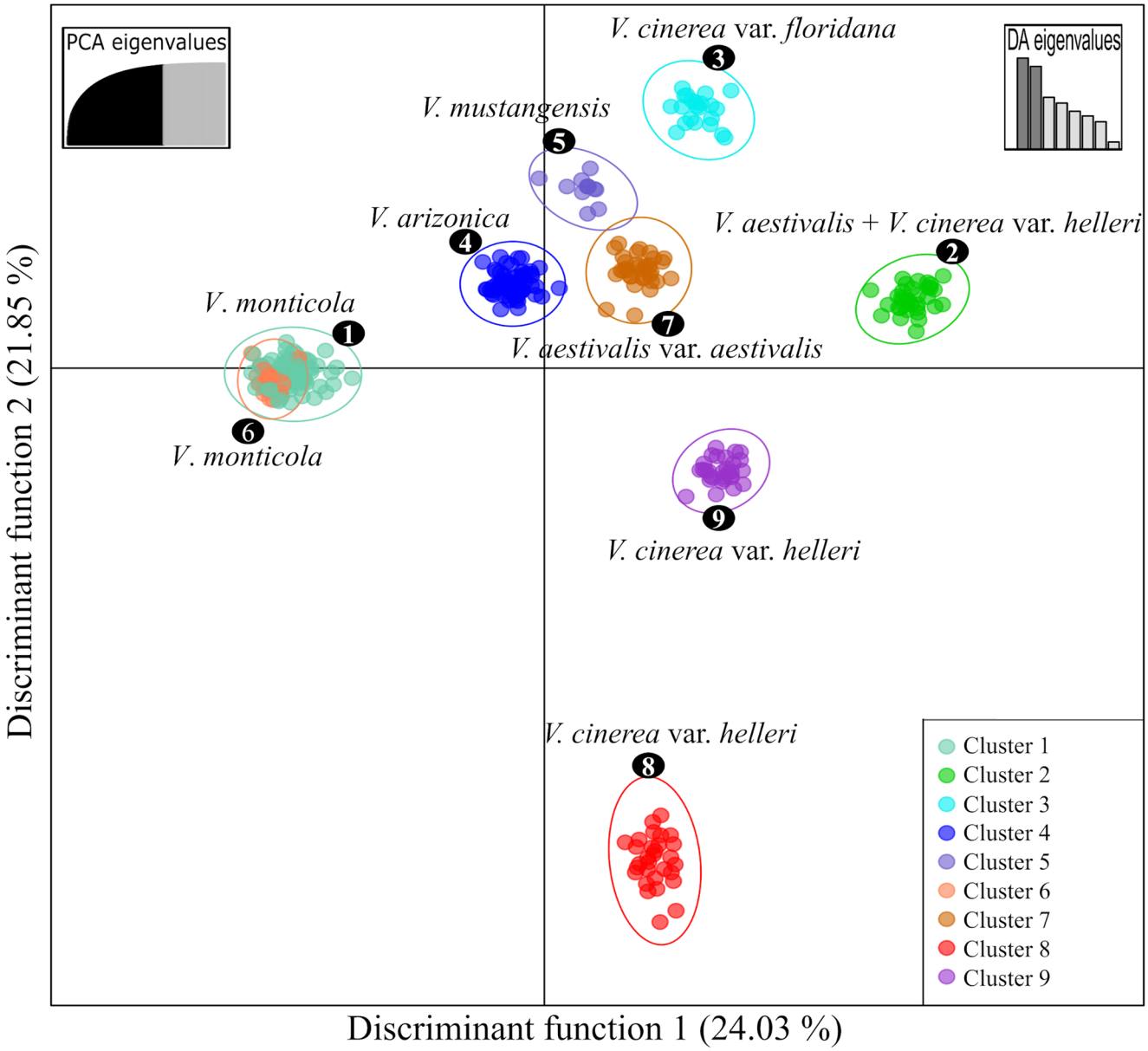

Fig. 4 DAPC scatterplots based on the K-means algorithm were used to determine the optimal number of clusters. Each dot represents an individual, with clusters represented in distinct colors. Accessions from different *Vitis* species were assigned to nine clusters (K=9): Clusters 1 and 6, predominantly *V. monticola* accessions; Cluster 2, composed of *V. aestivalis* and *V. cinerea* var. *helleri* accessions; Cluster 3, *V. cinerea* var. *floridana* accessions; Cluster 4, *V. arizonica* accessions; Cluster 5, *V. mustangensis* accessions; Cluster 7, predominantly *V. aestivalis* var. *aestivalis* accessions; and Clusters 8 and 9, associated with *V. cinerea* var. *helleri*

320) were placed with *V. cinerea* var. *floridana* accessions. The dendrogram also highlighted some outlier accessions, particularly those without a defined species and the remaining *V. rotundifolia* accessions. One of the unknown accessions, Arizonica Carlin-126 (ID 126), did not cluster with any recognized species and remained isolated. The other unknown accessions, along with the remaining *V. rotundifolia* accessions, were positioned near certain species groups but located on distinct branches, indicating potential taxonomic divergence.

3.3 Development of a core collection

Three independent sampling proportions, ranging from 10 to 30% of the total dataset, were established to minimize redundancy in the collection and enhance its practical application in crossbreeding programs. The objective was to identify the smallest subset of accessions capable of representing the available allelic diversity. Core10, composed of 32 accessions, captured 464 alleles, corresponding to 72.16% of the total (Table 4). Core20, with 62 accessions,

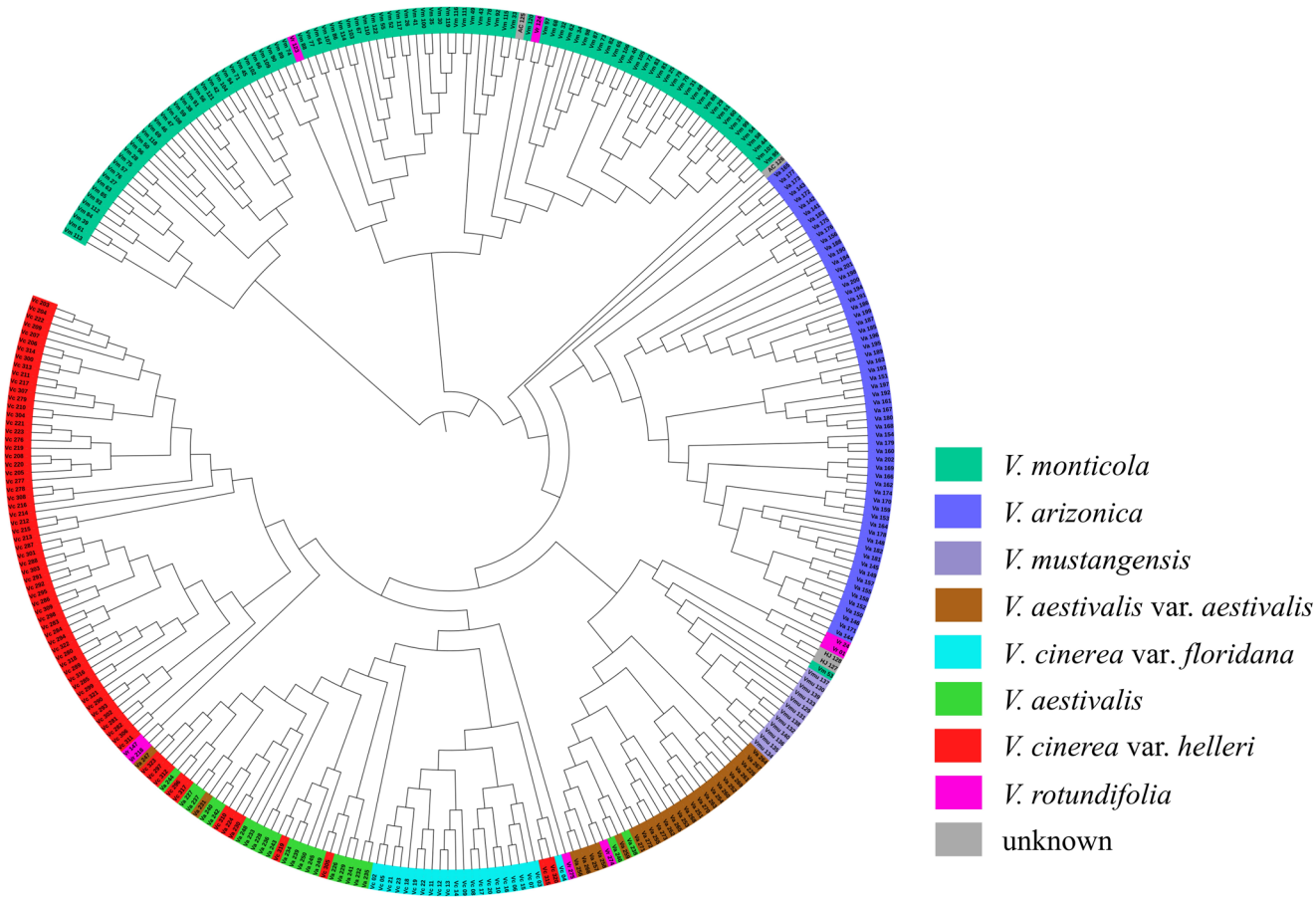

Fig. 5 Neighbor-joining dendrogram illustrating the genetic relationships among 323 *Vitis* accessions based on 29 SSR loci. Accessions are color-coded to represent their respective species groups

Table 4 Comparison of genetic diversity between the sampled core collections and the entire dataset. N: number of accessions. Na: number of alleles. Ne: number of effective alleles. H_O: observed heterozygosity. H_E: expected heterozygosity. I: shannon's information index

Sample name	*N*	Na	Ne	H_O*	H_E*	I*	Allele coverage (%)
Core10	32	464	273.64	0.69 (0.02)	0.86 (0.02)	2.36 (0.09)	72.16
Core20	62	583	276.25	0.73 (0.02)	0.86 (0.01)	2.45 (0.09)	90.66
Core30	95	643	283.32	0.74 (0.02)	0.86 (0.02)	2.47 (0.09)	**100**
Whole collection	323	643	260.15	0.74 (0.03)	0.86 (0.02)	2.39 (0.08)	**100**

*The standard error is shown in parentheses

encompassed 583 alleles (90.66%), whereas Core30, consisting of 95 accessions, successfully captured all 643 alleles present in the whole collection. The genetic diversity indices obtained for the core subsets were similar to or greater than those of the entire germplasm. The H_O values ranged from 0.69 (Core10) to 0.74 (Core30), with the latter being identical to that observed for all 323 accessions. The three subsets presented H_E values equal to those observed for the complete dataset (0.86). In contrast, the Ne values of the samples were higher than those of the entire collection, ranging from 273.64 (Core10) to 283.32 (Core30). Finally, the Shannon's information index (I) ranged from 2.36 (Core10) to 2.47 (Core30), which was lower than that of the whole collection only for Core10.

The Core30 sample, comprising 29.41% of the whole collection, was the only one capable of representing 100% of the alleles identified using 29 SSR markers across all 323 accessions. All clusters detected in the STRUCTURE and DAPC analyses are represented in Core30. Specifically, in the DAPC analysis (K = 9), 22 accessions were observed in Cluster 1, 14 in Cluster 2, 6 in Cluster 3, 13 in Cluster 4, 5 in Cluster 5, 3 in Cluster 6, 18 in Cluster 7, 5 in Cluster 8, and 9 in Cluster 9 (Fig. 6). When considering species-based grouping, the composition of Core30 was not exactly proportional to the total number of accessions in each group.

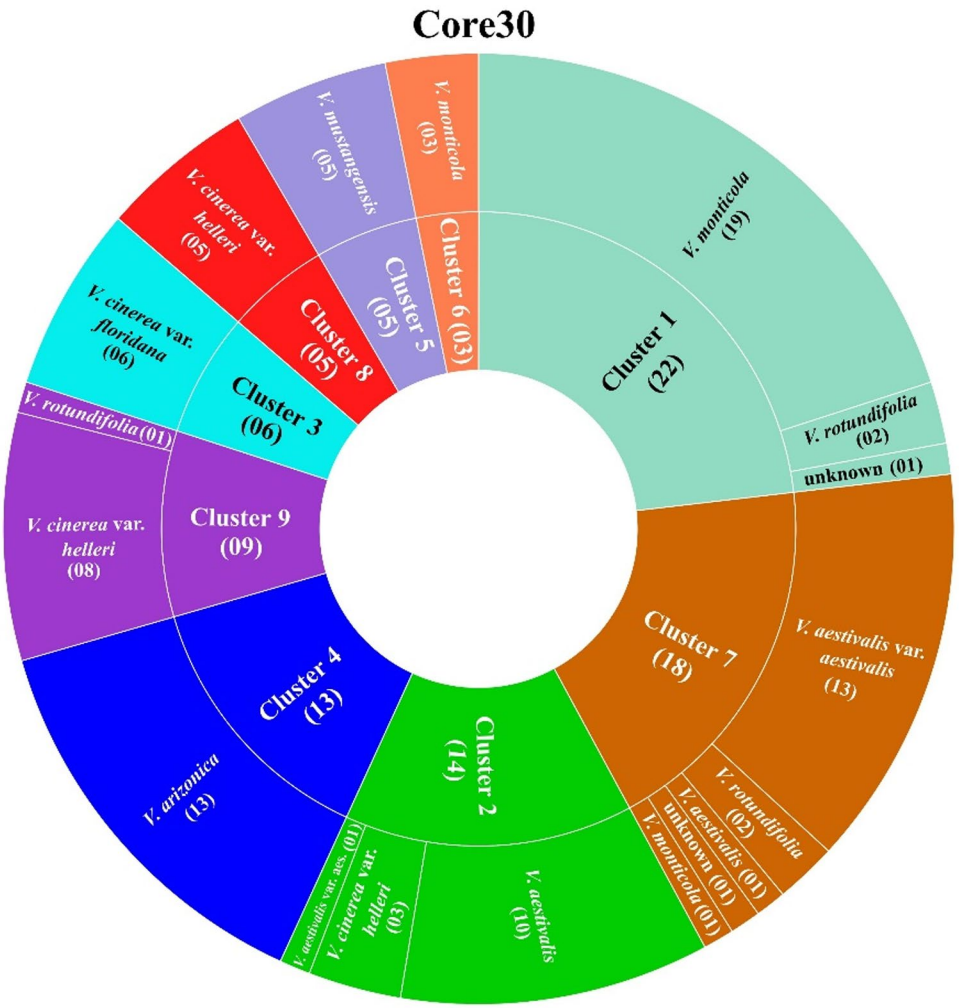

Fig. 6 Composition of the Core30 sample based on groups identified via the DAPC analysis, highlighting the distribution of species within each cluster. The colors used in the sunburst chart follow the same color scheme as the DAPC analysis for group representation

The core collection consisted of 23 accessions classified as *V. monticola*, 16 as *V. cinerea* var. *helleri*, 14 as *V. aestivalis* var. *aestivalis*, 13 as *V. arizonica*, 11 as *V. aestivalis*, 6 as *V. cinerea* var. *floridana*, 5 as *V. rotundifolia*, 5 as *V. mustangensis*, and 2 classified as unknown. Interestingly, almost all accessions of *V. rotundifolia* were included in the Core30 sample, whereas fewer *V. arizonica* accessions were selected compared to *V. aestivalis* var. *aestivalis*, despite *V. arizonica* exhibiting more than twice the total number of accessions.

3.4 R-loci detection

Fifteen screenable and reliable *R*-loci were examined in all 323 wild grapevine accessions, along with the reference genotype for each *R*-locus described in Table 1. The *Rpv3* locus represents a unique case in which different resistant haplotypes have been characterized (Di Gaspero et al. 2012); for this locus, we used a paired status (e.g., *Rpv3*-1 + *Rpv3*-2). A list of the *R*-loci detected in each genotype of the wild germplasm collection is shown in Supplementary Table S3. The *R*-loci *Rpv1*, *Rpv10*, *Rpv12*, *Ren1*, *Run1*, and *Rgb3* were not detected in any of the analyzed genotypes. Among the PM *R*-loci, *Ren3* was the most abundant (26.9%), followed by *Ren9* (5.5%) and *Ren2* (2.5%). While *Ren3* and *Ren9* were distributed across different species groups, *Ren2* was exclusively identified in eight *V. cinerea* var. *helleri* accessions within the VCH1 group (DAPC Cluster 8).

With respect to the DM *R*-loci, *Rpv3* was identified in 62 genotypes (19.19%). The *Rpv3*-2 haplotype was the most prevalent in the germplasm, found in 30 genotypes, followed by $Rpv3^{null-287}$ (9 genotypes), *Rpv3*-1 (8 genotypes), $Rpv3^{361-299}$ (7 genotypes), *Rpv3*-3 (6 genotypes), and $Rpv3^{321-312}$ (2 genotypes). Notably, the *Rpv3*-3 haplotype was detected exclusively in *V. aestivalis* accessions within the VAE group (DAPC Cluster 7), whereas $Rpv3^{321-312}$ was detected only in *V. arizonica* accessions from the VAR group (DAPC Cluster 4). The other DM *R*-locus, *Rpv27*, was identified in 19 genotypes (5.9%), including 4 classified as *V. monticola*, 13 as *V. aestivalis* var. *aestivalis*, and 3 as *V. aestivalis*. These genotypes were predominantly distributed across the VAE (DAPC Cluster 7) and VMO2 (DAPC

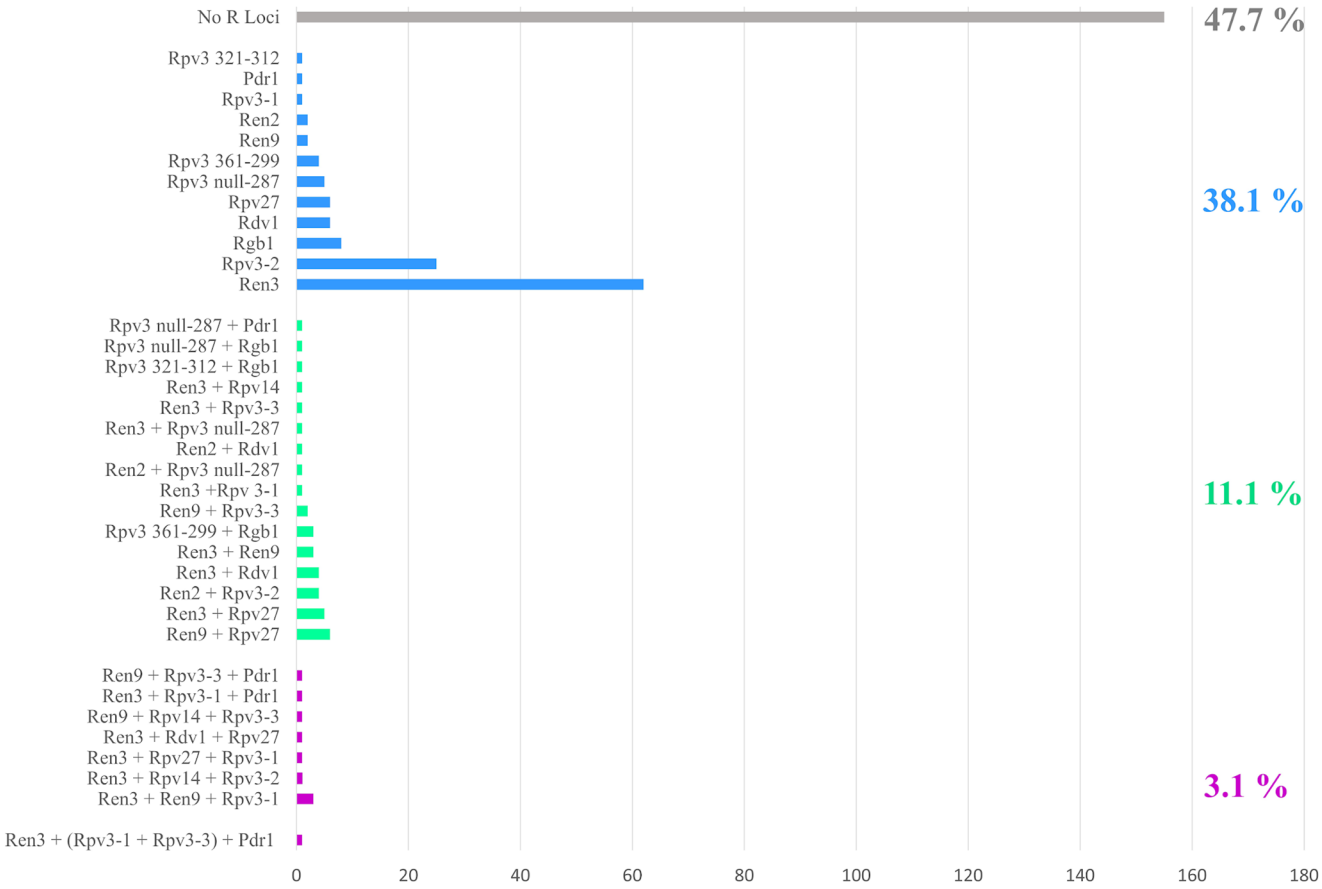

Fig. 7 Grouped bar chart showing the percentage of accessions with a single resistance (*R*) locus (blue), a combination of two (green) or three *R*-loci (magenta), and the absence of any investigated *R*-locus (gray). *Rpv*: resistance to *Plasmopara viticola*; *Run*: resistance to *Uncinula necator* (from *Muscadinia* spp.); *Ren*: resistance to *Erysiphe necator* (from *Vitis* spp.); *Rgb*: resistance to *Phyllosticta ampelicida*; *PdR*: resistance to *Xylella fastidiosa*; *Rdv*: resistance to *Daktulosphaira vitifoliae*

Cluster 1) groups. Additionally, *Rpv14* was detected in only 3 (0.9%) genotypes, comprising 2 classified as *V. aestivalis* from the VA-VC2 group (DAPC Cluster 2) and 1 as *V. aestivalis* var. *aestivalis* from the VAE group.

In terms of BR *R*-loci, *Rgb1* was detected in 13 genotypes (4%), with 7 classified as *V. monticola* and 6 as *V. arizonica*, primarily distributed within the VMO2 and VAR groups, respectively. Additionally, five genotypes (1.5%) harboring the *PdR1* locus associated with PD resistance were identified. Among these, two genotypes were classified as *V. arizonica* from the VAR group, and three were classified as *V. aestivalis* var. *aestivalis* from the VAE group. Finally, the *Rdv1* locus associated with phylloxera resistance was detected in 12 genotypes (3.7%), comprising 7 *V. monticola*, 4 *V. cinerea* var. *helleri*, and 1 *V. aestivalis* var. *aestivalis*, which are distributed across different genetic groups.

Notably, in 47.7% of the analyzed genotypes, none of the examined *R*-loci were detected (Fig. 7). Within the collection, 38.1% of the genotypes carried a single *R*-locus associated with resistance to at least one of the studied diseases, whereas 11.1% carried two loci. Furthermore, our findings revealed that pyramiding (stacking of at least two *R*-loci against the same disease) of *Rpv* loci occurred in only four (1.2%) genotypes; for *Run*/*Ren* loci, pyramiding was observed in six (1.8%) genotypes within the collection. Only ten individuals (3.1%) harbored three *R*-loci, including three occurrences in *V. monticola* featuring the combination *Ren3*+*Ren9*+*Rpv3*-1 and seven occurrences in *V. aestivalis* species, each exhibiting a distinct combination. Notably, one individual among the *V. aestivalis* var. *aestivalis* group (ID 263) exhibited a unique combination, possessing two variants of *Rpv3* (*Rpv3*-1 and *Rpv3*-3) in addition to *Ren3* and *PdR1*.

4 Discussion

4.1 Assessing genetic diversity

For advancements in sustainable grapevine crossbreeding programs, strategically utilizing genetic resources and thoroughly characterizing genetic diversity are crucial. This

strategy involves both the identification and incorporation of diverse genetic materials and an understanding of their genetic makeup and variability to enable more effective and targeted breeding efforts (Kaya et al. 2023; Magon et al. 2023). This study provides a valuable starting point for the introgression of North American wild *Vitis* genetic background into diverse genetic foregrounds.

The genotyping results confirmed that the collection is composed exclusively of unique genotypes, which is consistent with their origin from seeds and the outcrossing reproductive system that is typical of wild grape species (Riaz et al. 2018a; Cunha et al. 2020). The high heterozygosity values highlight the potential of these genetic materials as valuable sources of diversity. The H_E values observed in the FEM wild *Vitis* germplasm exceeded those reported for collections composed solely of *V. vinifera* accessions (e.g. Riaz et al. 2018a; Zdunić et al. 2020); however, they were comparable to values observed in non-*vinifera* collections (Migliaro et al. 2019; Riaz et al. 2020; de Oliveira et al. 2023). The taxonomic breadth of this germplasm is further evidenced by the high Na (~22), due to the fact that a relatively large proportion of the identified alleles were rare (36%). Overall, the wild grape germplasm represents a valuable reservoir of unique alleles, thus providing breeders with a diverse set of genetic resources that can support the development of stress-resilient cultivars (de Oliveira et al. 2020).

The extensive array of alleles generated by the 29 SSR primer set significantly influenced both the PIC and D*j*. In accordance with the categorization of Botstein et al. (1980), most of the SSR loci were classified as highly informative (PIC > 0.50), thereby confirming their suitability for assessing genetic diversity and structure. Only GF09-47 (0.32) exhibited reduced polymorphism and the lowest values of D*j*, Na, and Ne, potentially limiting its use in distinguishing closely related genotypes. The PIC is influenced by various factors, including the breeding behavior of the species, genetic diversity within the collection, collection size, and the genomic location targeted by the utilized primers in the study (Singh et al. 2013). A similar range of PIC values has been reported in other grapevine studies utilizing SSR markers (Marsal et al. 2016; Migliaro et al. 2019; Miazzi et al. 2020; de Oliveira et al. 2020).

Although null alleles were detected at four loci, the DAPC and dendrogram analyses that were performed after excluding these loci yielded clustering patterns and population structures that were highly consistent with those obtained by using the complete marker set (Supplementary Fig. S3). In a review of 233 studies, Dakin and Avise (2004) reported that the frequency of simulated null alleles is often overestimated, which may undervalue the informativeness of SSR markers. They also demonstrated that null alleles, at frequencies typically observed, are unlikely to introduce significant bias into assessments of genetic relationships. Additionally, Carlsson (2008) emphasized that an increase in the number of loci and the degree of genetic differentiation has a much greater impact on assignment accuracy than the presence of null alleles. This is particularly relevant in studies involving taxonomically broad collections or wild populations, where null alleles are more likely to occur. As long as the marker set retains sufficient polymorphism and discriminatory power, the overall reliability of population structure analyses remains unaffected. Nonetheless, loci with low null allele frequencies should always be preferred to minimize ambiguity and maximize assignment power (Dakin and Avise 2004; Carlsson 2008; Huang et al. 2022).

Given the lack of sampling information, such as the number of parent plants and seeds per plant collected to establish the germplasm, the diversity analysis among the species groups provided valuable insights into potential genetic bottlenecks and the levels of representativeness within each group. The analysis revealed a lower level of genetic variability within *V. mustangensis* than in the other groups (Table 3). The diversity parameters indicate a more homogeneous population, with the fixation of specific alleles. The smaller population size likely increased inbreeding, suggesting that the accessions from this species may have originated from a limited number of progenitors from a specific population. In the case of the *V. rotundifolia* species group, the analysis indicated high genetic variability ($H_E = 0.84$ and $I = 1.80$), despite the small number of accessions analyzed, suggesting the presence of highly divergent individuals within the group. However, signs of inbreeding were also observed within this species group ($H_E > H_O$ and $F = 0.21$), possibly resulting from the presence of genetically similar individuals with a shared origin, an effect that may be aggravated by the group's small size. The other species groups exhibited genetic diversity ranging from moderate to high, with evidence of substantial variation among individuals within each group, confirming that the accessions possess the genetic potential to integrate new materials into crossbreeding programs.

The presence of private alleles across all species groups underscores the importance of their conservation, as well as their potential as unique sources of genetic variation. These alleles can be strategically utilized in crossbreeding programs and genetic studies to increase allele richness, enhance adaptability and improve resistance to changing environmental conditions (de Souza et al. 2015; Salem and Sallam 2015; Sun et al. 2016) In this study, species groups with a greater number of individuals, such as *V. monticola* and *V. cinerea* var. *helleri*, exhibited a greater number of private alleles. This outcome may reflect a broader sampling of genetic variation; however, it could also indicate

greater intrinsic diversity, as the population size is not always directly correlated with the number of private alleles (de Souza et al. 2015; de Oliveira et al. 2020; Buck and Worthington 2022). Despite the small sample size, the *V. rotundifolia* group exhibited a high proportion of private alleles, thereby emphasizing its genetic distinctiveness.

4.2 Germplasm genetic structure and clustering

Without detailed information on the genetic relationships among most genotypes, as is the case in this study, identifying the most accurate clustering method is challenging. The application of multiple clustering criteria and the use of the predominant structure from these methods are essential to ensure that the results are not merely artifacts of the utilized techniques (Liang et al. 2015; Rodriguez et al. 2019; Tian et al. 2024). The consistency between the STRUCTURE, DAPC, and NJ dendrogram results reinforces the accuracy of the analyses and highlights a robust genetic structure within the evaluated germplasm, thereby supporting the hypothesis that dividing the germplasm into nine clusters is the most appropriate approach.

The identification of substructuring within the species *V. monticola*, *V. cinerea*, and *V. aestivalis*, which was highlighted in all three clustering analyses, provides additional insights into the origins of the accessions and helps identify varying levels of genetic relatedness among them within the species group. This divergence often arises from factors such as geographic isolation or the sampling of individuals from different locations along a geographical gradient (Maurya et al. 2024). In a previous study, Riaz et al. (2020) reported significant differences among *V. cinerea* accessions collected from different areas of North America, thus indicating regional types or varieties; moreover, they emphasized that species from different regions may contribute to unique genetic pools. Recently, Margaryan et al. (2023) also demonstrated that the wild population exhibits distinct grouping patterns and differentiation based on geographic location by analyzing the genetic structure of *V. vinifera* ssp. *sylvestris* in Armenia. In the case of the accessions used in this study, although the mother plants from which the seeds were sampled belonged to the same species, they may have originated from different populations, likely from geographically distinct regions. The *V. monticola* accessions classified as admixed via STRUCTURE may result from crosses between parent plants from different populations (such as VMO1 and VMO2), as these parents may be located close to each other in the grapevine collection field at the University of California, where the seeds were collected.

The relationships among the North American *Vitis* species analyzed in this study align with previous findings, which identified *V. monticola* and *V. arizonica* as the most genetically divergent species, belonging to distinct clades (Wan et al. 2013; Klein et al. 2018; Zecca et al. 2020). Conversely, these studies also revealed that *V. mustangensis*, *V. cinerea*, and *V. aestivalis* share a closer genetic relationship, with their clusters branching from a common node. According to a study conducted by Klein et al. (2018), *V. cinerea* and *V. aestivalis* exhibit even greater genetic proximity and are part of the same subclade. This close relationship may explain the clustering of certain *V. aestivalis* and *V. cinerea* accessions in the first round of the STRUCTURE analysis (VA-VC group). In the DAPC analysis and second round of the STRUCTURE analysis, the *V. cinerea* var. *floridana* accessions formed a distinct group (VCF group and DAPC Cluster 3). However, the clustering of *V. cinerea* var. *helleri* and *V. aestivalis* accessions remained evident (VA-VC2 group and DAPC Cluster 2) and was also observed in the NJ dendrogram. An alternative explanation could be the potential misclassification of some accessions, which might represent variants of either *V. cinerea* or *V. aestivalis*. In the study by Wan et al. (2013), accessions classified as *V. cinerea* var. *floridana* were consistently placed within the *V. aestivalis* clade. According to the authors, this classification highlights the historical confusion regarding the synonym *Vitis simpsonii*, which has been attributed to two different species, one belonging to the Aestivales group and the other to the Cinerescentes group. Studies conducted by Zecca et al. (2020) and Péros et al. (2023) revealed a closer genetic relationship between *V. cinerea* var. *floridana* and *V. aestivalis* than between *V. cinerea* var. *floridana* and other varieties of *V. cinerea*. Furthermore, Péros et al. (2023) suggested that *V. cinerea* var. *floridana* does not belong to *V. cinerea* but is more likely a pure species or a hybrid closely related to *V. aestivalis*. Zecca et al. (2020) further suggested that the placement of *V. cinerea* var. *floridana* within the "*V. aestivalis*-like" clade might indicate instances of cross-breeding where the ranges of these species overlap. These findings support the hypothesis that the accessions in the VA-VC2 group (DAPC Cluster 2), due to their genetic proximity to the *V. cinerea* var. *floridana* cluster, are actually all *V. aestivalis*-like or hybrids closely related to this species, with misidentification being limited to the accessions in this group labeled *V. cinerea* var. *helleri*.

Passport data for wild species often contain incomplete or inaccurate information. Therefore, molecular data are frequently used to identify genetic signatures that aid in classifying genotypes within species or to flag accession identities as suspects (Klein et al. 2018; Riaz et al. 2020; Zecca et al. 2020). In this study, the NJ dendrogram suggested taxon names for one previously unknown accession, whereas nine accessions were potentially misidentified in the FEM wild *Vitis* germplasm. Although the NJ dendrogram analysis could not definitively place some accessions

within the clades due to low bootstrap support, it provided indications of their close association with certain species groups (Klein et al. 2018). The DAPC and STRUCTURE analyses support the reidentification of most of these accessions; however, morphological analyses should be conducted for more accurate conclusions. The lack of grouping among the *V. rotundifolia* accessions reinforces the possibility of misidentification and supports the findings from the diversity analysis of this species group, which revealed the presence of highly divergent individuals, as reflected in the high H_E value. Additionally, some accessions were closely clustered in the dendrogram, suggesting the presence of related individuals within the species group, which aligns with the inbreeding signals identified in the diversity analysis ($H_E > H_O$ and $F = 0.21$).

4.3 Building the core collection

The aim of developing a core collection is to increase the use, management, and conservation of germplasms by providing a subset of accessions that captures the majority of the genetic diversity within the entire collection while minimizing redundancy (de Souza et al. 2015; Anglin et al. 2024). The size of a core collection is influenced by multiple factors, including the heterozygosity of the species, the germplasm population size, the level of genetic diversity, and the sampling strategy utilized (Wang et al. 2023). In this study, approximately 30% of the accessions (Core30) were sampled to capture the entire allelic diversity present in the whole collection. In other collections consisting solely of the species *V. vinifera*, core collections with smaller percentages of individuals (4–15%) have been reported (Cunff et al. 2008; Cipriani et al. 2010; Emanuelli et al. 2013). Conversely, in taxonomically broader collections that include wild species and interspecific hybrids, the percentage of individuals tends to be similar to that observed in this study (~30%) (Upadhyay et al. 2013; Migliaro et al. 2019; de Oliveira et al. 2020). This result is due to the fact that cultivated genotypes tend to be less diverse compared to their wild counterparts (Cunff et al. 2008; Laucou et al. 2011). Consequently, collections that include wild accessions from different species require a larger number of individuals to adequately represent genetic diversity, which is due to the higher frequency of rare alleles (Lamboy and Alpha 1998), many of which are unique to specific species, as evidenced in the present study.

The designated core collection (Core30) provided a fair and proportionally balanced representation of each of the nine groups that were identified in the structure analyses, with a composition proportional to the number of accessions belonging to each species group. The genetic diversity indices obtained for the core collection were similar to those of the original collection, and all of the alleles present in the original collection were preserved. These results indicate that the core collection maintains sufficient genetic variation, thus making it a strategic starting point for the incorporation of this genetic resource into molecular breeding programs. However, although all of the alleles detected by the markers utilized in this study are fully represented in the core collection, this scenario does not necessarily ensure the inclusion of the full spectrum of functional genetic diversity present in the entire germplasm, due to the fact that the preservation of allelic diversity is not always compatible with the preservation of phenotypic variability (Mahmoodi et al. 2021; Xue et al. 2021). Moreover, the utilized SSR markers are neutral and not gene-based; thus, they primarily reflect polymorphisms in non-coding DNA regions and may not capture the variation that is directly associated with functional traits. Future efforts to refine the core collection may benefit from the integration of additional layers of characterization, including functional markers, phenotypic evaluations, and ecogeographical traits, in order to further enhance its functional representativeness and maximize its utility for breeding programs (Kumar et al. 2016; Boccacci et al. 2021).

4.4 Identifying potential sources of disease and pest resistance

Advancements in genetics/genomics and crossbreeding applications have utilized wild grapevine plants to introduce disease resistance genes and increase tolerance to a variety of abiotic stress factors, as well as diverse climatic conditions (Vezzulli et al. 2022). The identification of genetic markers associated with agronomic and fruit quality traits would enable the screening of large plant populations and progenies, allowing the selection of individuals carrying beneficial alleles (De Lorenzis et al. 2022). Moreover, not all of the identified *R*-loci have suitable marker choices available in the literature. In this study, we utilized only *R*-loci associated with resistance to different pathogens that have a defined set of robust and exploitable SSR markers for MAS. The lack of detection of the *R*-loci *Rpv1*, *Rpv10*, *Rpv12*, *Ren1*, and *Run1* in any of the analyzed genotypes can be attributed to the absence of genotypes derived from *V. amurensis* and *V. vinifera* species, as well as to the low representation of *V. rotundifolia* species in the collection. However, alternative technical factors should also be considered, as failures in the amplification of one of the two alleles at a target locus in a competitive PCR reaction may result from either sequence-independent factors or allele-specific sequence variations (Blais et al. 2015). Mutations at one or both primer-binding sites can lead to mismatches between the primer and the target DNA, thereby potentially

resulting in allele dropout and misclassification of heterozygous individuals as being homozygous. Moreover, the phenomenon of allelic dropout may also involve preferential amplification of one allele over the other during the PCR thermocycling process, which may occur due to stochastic effects or technical issues such as suboptimal DNA quality, PCR inhibitors, or variations in thermocycling conditions (Guichoux et al. 2011; Blais et al. 2015; Jahnke et al. 2022). Therefore, the absence of these *R*-loci must be interpreted with caution, as technical limitations in molecular detection cannot be entirely ruled out.

Interestingly, nearly half of the genotypes lacked any of the analyzed *R*-loci associated with resistance to the investigated biotic factors, including all of the *V. mustangensis* accessions. However, this absence does not necessarily imply susceptibility; rather, it highlights an opportunity to identify novel resistance sources. Although further analyses are required for definitive conclusions, ongoing studies, including a preliminary phenotypic screening for DM resistance in greenhouse-grown potted plants, have indicated the presence of resistant genotypes among those lacking detectable *Rpv* loci in this study. In this experiment, eight accessions were randomly selected and inoculated following the protocol described by Vezzulli et al. (2018) (additional details are provided in Supplementary Table S4). Among them, two accessions lacking any of the evaluated *R*-loci exhibited high resistance to DM, based on the OIV descriptor 452 (OIV 2022), including one *V. mustangensis* accession, which may represent a potential new donor.

When considering the occurrence of each *R*-locus independently, regardless of its combination with other *R*-loci, a higher frequency of *Ren3* and *Rpv3* was noted. This finding is consistent with the results reported by Zini et al. (2019) in a collection of hybrids, where a significant predominance of *Rpv3* and *Ren3* was also observed. The resistance to *E. necator* conferred by the *Ren3* and *Ren9* loci is believed to stem from an as-yet-unidentified North American *Vitis* species (Zendler et al. 2017). In this study, these two loci in different species were grouped into distinct genetic pools via structural analyses. Additionally, some loci, such as the *Rpv3*, *Rpv14*, *Rpv27*, *Rdv1*, and *PdR1* loci, have been identified in species other than those described as resistance sources. Different haplotypes of an *R*-locus can emerge randomly and be preserved in isolated populations with varying genetic backgrounds if they provide resistance to local pathogen variants. These haplotypes lead to varying levels of field resistance (Foria et al. 2018). However, exploratory studies of *R*-loci across different genetic backgrounds from the discovered resistance source are rare. One hypothesis for the observed results is that related species sharing similar environments exposed to the same selection pressure may share a defense mechanism against the same pathogen as a form of parallel evolution (Bailey et al. 2017). However, SSRs have limitations that can lead to misleading interpretations for MAS, such as the occurrence of homoplasy, where two alleles are identical in state but not identical by descent. This occurrence results in instances of evolutionarily independent events that can confound studies of genetic variation within and among populations (Viard et al. 1998). Regardless, the inclusion of field resistance data will be crucial for elucidating these hypotheses, assessing the level of resistance in the field, and validating the MAS results.

In this study, six distinct *Rpv3* haplotypes were identified among the accessions classified as *V. aestivalis*, *V. cinerea*, *V. arizonica*, and *V. monticola*. Previous analyses of the *Rpv3* locus in North American *Vitis* species revealed seven conserved haplotypes exclusive to resistant accessions (Di Gaspero et al. 2012), thereby suggesting that this locus in resistant breeding lines originated from multiple ancestral sources. Among the haplotypes detected in this study, *Rpv3-2* was the most prevalent. However, its association with resistance depends on the presence of at least one null allele at the UDV305 locus. The detection of this null allele complicates MAS in genotypes lacking documented pedigrees, as it is not possible to determine whether the signal reflects a true null allele or a homozygous state for alleles of identical size. Similar limitations apply to the *Rpv3-3* and *Rpv3*$^{\text{null}-287}$ haplotypes, thus underscoring the importance of incorporating phenotypic data to validate resistance predictions, particularly in wild populations. To sum up, in addition to the multihaplotypic nature of the *Rpv3* locus, it is important to consider that a recent seminal study demonstrated that the same haplotype can underlie (slightly) different resistance genes, which may impact the actual host-pathogen interaction (Wilkerson et al. 2025).

Different North American *Vitis* species, including *V. monticola*, which possesses seven genotypes of the *Rgb1* locus identified in this study, have been reported to exhibit high levels of resistance to BR (Bettinelli et al. 2023b). In contrast, *V. arizonica*, which also carried *Rgb1* genotypes in this collection, has been described as susceptible in a recent review (Bettinelli et al. 2023b). This susceptibility may result from pathogen strains capable of overcoming specific resistance mechanisms, which is possibly due to selective pressures from prolonged coexistence with resistant hosts. Additionally, differences in pathogen strain composition across years or regions, as well as the inherent genetic variability within species, may influence resistance outcomes, and these factors are difficult to identify in studies with limited sampling (Bettinelli et al. 2023b). Therefore, caution is needed when generalizing susceptibility to an entire species. Notably, this study reports for the first time the presence

of *Rgb1* in six *V. arizonica* genotypes, thereby highlighting their potential as sources of BR resistance.

The *Rdv1* locus, associated with phylloxera resistance via a hypersensitive response, is thought to have originated from *V. cinerea* (Zhang et al. 2009; Hausmann et al. 2011). Among the 12 identified accessions carrying *Rdv1*, four were classified as *V. cinerea*, and seven were classified as *V. monticola*. Although phylloxera resistance in *V. monticola* has been rarely reported, this species was previously classified as tolerant (Wapshere and Helm 1987). Its limited use in rootstock breeding is likely due to its poor rooting ability and very slow growth (Cousins and Lauver 2003; Heinitz et al. 2019). However, the results of the present study, which demonstrated the presence of multiple *R*-loci conferring resistance to PM, DM, BR, and phylloxera in *V. monticola*, underscore its potential as a valuable genetic resource for biotic stress resistance. Notably, species such as *V. monticola* and *V. cinerea* var. *helleri* are currently restricted to narrow natural habitats and experience significant threats from human activities and invasive species (Heinitz et al. 2019; Zecca et al. 2020). The survival and establishment of the accessions analyzed in this study under the harsh winter conditions of northern Italy further demonstrate their adaptability and potential utility in European grapevine crossbreeding programs.

The *PdR1* locus, mapped to chromosome 14, has played a key role in developing PD-resistant grapevine varieties (Krivanek et al. 2006; Riaz et al. 2009). In this study, five accessions carrying the *PdR1* locus were identified, belonging to *V. arizonica* and *V. aestivalis* var. *aestivalis*. Previous studies have also reported *PdR1* in *V. arizonica* accessions resistant to PD, including both pure species and hybrids distributed across the arid southwestern USA and northern Mexico (Krivanek et al. 2005; Riaz et al. 2006, 2018b, 2023). Furthermore, Riaz et al. (2020) demonstrated that PD resistance extends beyond *V. arizonica*, as other North American wild *Vitis* species have demonstrated varying levels of resistance to the disease. Notably, accessions of *V. aestivalis* have exhibited the ability to tolerate relatively high bacterial loads without excessive leaf scorch. Moreover, *V. aestivalis* has been incorporated into the genetic background of several PD-resistant varieties released in the USA (Riaz et al. 2020). In Europe, Pierce's Disease has been observed as an emerging threat since its first official detection in grapevines in Mallorca (Balearic Islands) in 2017 (Moralejo et al. 2019). Thus, the establishment of effective control strategies and the initiation of molecular breeding programs aimed at developing resistant cultivars are urgently needed. The results obtained in this study have facilitated the identification of potential sources of PD resistance, thereby providing valuable genetic combinations for upcoming crossbreeding efforts.

5 Conclusions

The molecular characterization of germplasm collections is essential for the efficient management and utilization of genetic resources. In this study, the comprehensive genotyping of the FEM wild grapevine collection revealed high levels of genetic diversity and the presence of key *R*-loci, thus underscoring the importance of the conservation of this germplasm both for its ecological value and for its potential to contribute to sustainable viticulture. These genotypes represent valuable sources of known resistance alleles and offer opportunities to identify novel resistance traits. However, the limited number of available reliable markers compared to the numerous QTLs described in grapevine research highlights the need to integrate phenotypic data. Such integration is crucial to validate the MAS results, accurately assess resistance levels, and validate new *R*-loci. The structure and genetic diversity analyses provided important insights into the relationships among the genotypes, thereby enabling the identification of distinct genetic pools within species groups. This information is especially valuable for crossbreeding programs, given the limitations of passport and sampling data associated with these accessions. Furthermore, the establishment of a core collection offers a practical approach for the targeted identification of candidate genes as well as the evaluation of key traits, thus reducing redundancy while ensuring representative genetic diversity. Overall, this study represents a milestone for future molecular breeding strategies, boosting marker- and genomics-assisted breeding and potentially supporting the application of new genomic techniques applications.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13580-025-00749-0.

Acknowledgements The authors are grateful to the personnel of the Grapevine Genetics and Breeding Unit (FEM) for their greenhouse and field support. In addition, they would like to thank the Università degli Studi di Milano (UniMI) for the donation of the wild genetic material.

Author contributions GLO - conceptualization, data observation, data processing, formal analysis, manuscript writing, and paper editing. PB - formal analysis and paper editing. DN - data observation and data processing. APS - funding acquisition and paper editing. MFMF - review and paper editing. MS - funding acquisition and project administration. SV - development of the idea, design of the experiment, overall supervision of the investigation, funding acquisition and paper editing. All of the authors have critically revised this work and have approved the submitted version.

Funding Open access funding provided by Fondazione Edmund Mach - Istituto Agrario di San Michele all'Adige within the CRUI-CARE Agreement. The Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) provided a fellowship to GLO (200325/2022-1). The study was partially supported by the Agritech National Research

Center and received funding from the European Union Next-Generation EU (Piano Nazionale di Ripresa e Resilienza (PNRR)—Missione 4 Componente 2, Investimento 1.4—D.D. 1032 17/06/2022, CN00000022).

Data availability All of the data supporting the conclusions of this study are included in this article.

Declarations

Ethical approval This article does not contain any experiments with human participants or animals performed by any of the authors.

Competing interests The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

Aguirre-Liguori JA, Morales-Cruz A, Gaut BS (2022) Evaluating the persistence and utility of five wild vitis species in the context of climate change. Mol Ecol 31:6457–6472. https://doi.org/10.1111/mec.16715

Amiteye S (2021) Basic concepts and methodologies of DNA marker systems in plant molecular breeding. Heliyon 7:e08093. https://doi.org/10.1016/j.heliyon.2021.e08093

Anglin NL, Chavez O, Soto - Torres J et al (2024) Promiscuous potato: elucidating genetic identity and the complex genetic relationships of a cultivated potato germplasm collection. Front Plant Sci 15. https://doi.org/10.3389/fpls.2024.1341788

Atak A (2024) Vitis species for stress tolerance/resistance. Genet Resour Crop Evol. https://doi.org/10.1007/s10722-024-02106-z

Bailey SF, Blanquart F, Bataillon T, Kassen R (2017) What drives parallel evolution? BioEssays. 39:1–9. https://doi.org/10.1002/bies.201600176

Bettinelli P, Nicolini D, Costantini L et al (2023a) Towards marker-assisted breeding for black rot bunch resistance: identification of a major QTL in the grapevine cultivar 'merzling'. Int J Mol Sci 24. https://doi.org/10.3390/ijms24043568

Bettinelli P, Nicolini D, Giovannini O et al (2023b) Breeding for black rot resistance in grapevine: advanced approaches for germplasm screening. Euphytica 219:1–16. https://doi.org/10.1007/s10681-023-03235-9

Blais J, Lavoie SB, Giroux S et al (2015) Risk of misdiagnosis due to allele dropout and false-positive PCR artifacts in molecular diagnostics. J Mol Diagnostics 17:505–514. https://doi.org/10.1016/j.jmoldx.2015.04.004

Boccacci P, Aramini M, Ordidge M et al (2021) Comparison of selection methods for the establishment of a core collection using SSR markers for hazelnut (Corylus Avellana L.) accessions from European germplasm repositories. Tree Genet Genomes 17. https://doi.org/10.1007/s11295-021-01526-7

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Bowers JE, Dangl GS, Vignani R, Meredith CP (1996) Isolation and characterization of new polymorphic simple sequence repeat loci in grape (Vitis vinifera L). Genome 39:628–633. https://doi.org/10.1139/g96-080

Bowers JE, Dangl GS, Meredith CP (1999) Development and characterization of additional microsatellite DNA markers for grape. Am J Enol Vitic 50:243–246

Buck K, Worthington M (2022) Genetic diversity of wild and cultivated muscadine grapes (Vitis rotundifolia Michx). Front Plant Sci 13. https://doi.org/10.3389/fpls.2022.852130

Carlsson J (2008) Effects of microsatellite null alleles on assignment testing. J Hered 99:616–623. https://doi.org/10.1093/jhered/esn048

Cipriani G, Spadotto A, Jurman I et al (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. Theor Appl Genet 121:1569–1585. https://doi.org/10.1007/s00122-010-1411-9

Coleman C, Copetti D, Cipriani G et al (2009) The powdery mildew resistance gene REN1 co-segregates with an NBS-LRR gene cluster in two central Asian grapevines. BMC Genet 10:1–20. https://doi.org/10.1186/1471-2156-10-89

Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc B Biol Sci 363:557–572. https://doi.org/10.1098/rstb.2007.2170

Cousins P, Lauver M (2003) Response to root-knot nematodes in a Vitis monticola hybrid population. Acta Hortic 319–324. https://doi.org/10.17660/ActaHortic.2003.623.36

Cunha J, Ibáñez J, Teixeira-Santos M et al (2020) Genetic relationships among Portuguese cultivated and wild Vitis vinifera L. Germplasm. Front Plant Sci 11:1–14. https://doi.org/10.3389/fpls.2020.00127

Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. Heredity (Edinb) 93:504–509. https://doi.org/10.1038/sj.hdy.6800545

De Beukelaer H, Davenport GF, Fack V (2018) Core Hunter 3: flexible core subset selection. BMC Bioinformatics 19:203. https://doi.org/10.1186/s12859-018-2209-z

De Lorenzis G, Carbonell-Bejerano P, Toffolatti SL, Tello J (2022) Editorial: advances in grapevine genetic improvement: towards high quality, sustainable grape production. Front Plant Sci 13. https://doi.org/10.3389/fpls.2022.1080733

de Oliveira GL, de Souza AP, de Oliveira FA et al (2020) Genetic structure and molecular diversity of Brazilian grapevine germplasm: management and use in breeding programs. PLoS One 15:e0240665. https://doi.org/10.1371/journal.pone.0240665

de Oliveira GL, Niederauer GF, de Oliveira FA et al (2023) Genetic diversity, population structure and parentage analysis of Brazilian grapevine hybrids after half a century of genetic breeding. Sci Hortic (Amsterdam) 311. https://doi.org/10.1016/j.scienta.2023.111825

de Souza LM, Le Guen V, Cerqueira-Silva CBM et al (2015) Genetic diversity strategy for the management and use of rubber genetic resources: more than 1,000 wild and cultivated accessions in a 100-genotype core collection. PLoS One 10:e0134607. https://doi.org/10.1371/journal.pone.0134607

Di Gaspero G, Cipriani G, Marrazzo MT et al (2005) Isolation of (AC)n-microsatellites in vitis vinifera L. and analysis of genetic

background in grapevines under marker assisted selection. Mol Breed 15:11–20. https://doi.org/10.1007/s11032-004-1362-4
Di Gaspero G, Copetti D, Coleman C et al (2012) Selective sweep at the Rpv3 locus during grapevine breeding for downy mildew resistance. Theor Appl Genet 124:277–286. https://doi.org/10.1007/s00122-011-1703-8
Doyle J (1991) DNA protocols for plants. Mol Tech Taxon 283–293. https://doi.org/10.1007/978-3-642-83962-7_18
Dry I, Riaz S, Fuchs M et al (2019) Scion breeding for resistance to biotic stresses. In: Cantu D, Walker MA (eds) The grape genome. Springer International Publishing, Cham, pp 319–347. https://doi.org/10.1007/978-3-030-18601-2_15
Earl DA, vonHoldt BM (2012) Structure harvester: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361. https://doi.org/10.1007/s12686-011-9548-7
Eibach R, Töpfer R (2015) Traditional grapevine breeding techniques. In: Grapevine breeding programs for the wine industry. Elsevier, pp 3–22. https://doi.org/10.1016/B978-1-78242-075-0.00001-6
Emanuelli F, Lorenzi S, Grzeskowiak L et al (2013) Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. BMC Plant Biol 13:39. https://doi.org/10.1186/1471-2229-13-39
Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x
Fechter I, Hausmann L, Zyprian E et al (2014) QTL analysis of flowering time and ripening traits suggests an impact of a genomic region on linkage group 1 in Vitis. Theor Appl Genet 127:1857–1872. https://doi.org/10.1007/s00122-014-2310-2
Foria S, Magris G, Morgante M, Di Gaspero G (2018) The genetic background modulates the intensity of Rpv3-dependent downy mildew resistance in grapevine. Plant Breed 137:220–228. https://doi.org/10.1111/pbr.12564
Goto-Yamamoto N, Mouri H, Azumi M, Edwards KJ (2006) Development of grape microsatellite markers and microsatellite analysis including oriental cultivars. Am J Enol Vitic 57:105–108. https://doi.org/10.5344/ajev.2006.57.1.105
Guichoux E, Lagache L, Wagner S et al (2011) Current trends in microsatellite genotyping. Mol Ecol Resour 11:591–611. https://doi.org/10.1111/j.1755-0998.2011.03014.x
Hausmann L, Eibach R, Zyprian E, Töpfer R (2011) Genetic analysis of phylloxera root resistance in cultivar 'Börner'. Acta Hortic 47–52. https://doi.org/10.17660/ActaHortic.2011.904.6
Heinitz CC, Uretsky J, Dodson Peterson JC et al (2019) Crop wild relatives of grape (Vitis vinifera L.) throughout North America. In: Greene SL, Williams KA, Khoury CK et al (eds) North American crop wild relatives, vol 2. Springer International Publishing, Cham, pp 329–351. https://doi.org/10.1007/978-3-319-97121-6_10
Huang C-J, Chu F-H, Huang Y-S et al (2022) SSR individual identification system construction and population genetics analysis for Chamaecyparis formosensis. Sci Rep 12:4126. https://doi.org/10.1038/s41598-022-07870-5
Jahnke G, Smidla J, Deák T et al (2022) The SSR null allele problem, and its consequences in pedigree reconstruction and population genetic studies in viticulture. Horticulturae 8:658. https://doi.org/10.3390/horticulturae8070658
Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806. https://doi.org/10.1093/bioinformatics/btm233
Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11. https://doi.org/10.1186/1471-2156-11-94
Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x
Kamvar ZN, Tabima JF, Gŕunwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2014:1–14. https://doi.org/10.7717/peerj.281
Kaya HB, Dilli Y, Oncu-Oner T, Ünal A (2023) Exploring genetic diversity and population structure of a large grapevine (Vitis vinifera L.) germplasm collection in Türkiye. Front Plant Sci 14:1–16. https://doi.org/10.3389/fpls.2023.1121811
Klein LL, Miller AJ, Ciotir C et al (2018) High-throughput sequencing data clarify evolutionary relationships among North American Vitis species and improve identification in USDA vitis germplasm collections. Am J Bot 105:215–226. https://doi.org/10.1002/ajb2.1033
Krivanek AF, Famula TR, Tenscher A, Walker MA (2005) Inheritance of resistance to xylella fastidiosa within a Vitis rupestris × Vitis arizonica hybrid population. Theor Appl Genet 111:110–119. https://doi.org/10.1007/s00122-005-1999-3
Krivanek AF, Riaz S, Walker MA (2006) Identification and molecular mapping of PdR1, a primary resistance gene to pierce's disease in vitis. Theor Appl Genet 112:1125–1131. https://doi.org/10.1007/s00122-006-0214-5
Kumar S, Ambreen H, Variath MT et al (2016) Utilization of molecular, phenotypic, and geographical diversity to develop compact composite core collection in the oilseed crop, safflower (Carthamus tinctorius L.) through maximization strategy. Front Plant Sci 7. https://doi.org/10.3389/fpls.2016.01554
Lamboy WF, Alpha CG (1998) Using simple sequence repeats (SSRs) for DNA fingerprinting germplasm accessions of grape (Vitis L.) species. J Am Soc Hortic Sci 123:182–188. https://doi.org/10.21273/JASHS.123.2.182
Laucou V, Lacombe T, Dechesne F et al (2011) High throughput analysis of grape genetic diversity as a tool for germplasm collection management. Theor Appl Genet 122:1233–1245. https://doi.org/10.1007/s00122-010-1527-y
Le Cunff L, Fournier-Level A, Laucou V et al (2008) Construction of nested genetic core collections to optimize the exploitation of natural diversity in Vitis vinifera L. subsp sativa. BMC Plant Biol 8:31. https://doi.org/10.1186/1471-2229-8-31
Letunic I, Bork P (2021) Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296. https://doi.org/10.1093/nar/gkab301
Liang W, Dondini L, De Franceschi P et al (2015) Genetic diversity, population structure and construction of a core collection of Apple cultivars from Italian germplasm. Plant Mol Biol Rep 33:458–473. https://doi.org/10.1007/s11105-014-0754-9
Ma Z-Y, Wen J, Ickert-Bond SM et al (2018) Phylogenomics, biogeography, and adaptive radiation of grapes. Mol Phylogenet Evol 129:258–267. https://doi.org/10.1016/j.ympev.2018.08.021
Magon G, De Rosa V, Martina M et al (2023) Boosting grapevine breeding for climate-smart viticulture: from genetic resources to predictive genomics. Front Plant Sci 14:1–15. https://doi.org/10.3389/fpls.2023.1293186
Mahmoodi R, Dadpour MR, Hassani D et al (2021) Composite core set construction and diversity analysis of Iranian walnut germplasm using molecular markers and phenotypic traits. PLoS One 16:e0248623. https://doi.org/10.1371/journal.pone.0248623
Margaryan K, Töpfer R, Gasparyan B et al (2023) Wild grapes of armenia: unexplored source of genetic diversity and disease resistance. Front Plant Sci 14. https://doi.org/10.3389/fpls.2023.1276764

Marsal G, Mateo-Sanz JM, Canals JM et al (2016) SSR analysis of 338 accessions planted in Penedès (Spain) reveals 28 unreported molecular profiles of Vitis vinifera L. Am J Enol Vitic 67:466–470. https://doi.org/10.5344/ajev.2016.16013

Maul E, Sudharma KN, Kecke S et al (2012) The European Vitis database (http://www.eu-vitis.de/index.php) - A technical innovation through an online uploading and interactive modification system E. Vitis - J Grapevine Res 51:79–85

Maurya R, Singh S, Babu YS et al (2024) Molecular diversity studies and core development in Sesame Germplasm (Sesamum indicum L.) using SSR markers. Plant Mol Biol Report. https://doi.org/10.1007/s11105-024-01476-z

Merdinoglu D, Butterlin G, Bevilacqua L et al (2005) Development and characterization of a large set of microsatellite markers in grapevine (Vitis vinifera L.) suitable for multiplex PCR. Mol Breed 15:349–366. https://doi.org/10.1007/s11032-004-7651-0

Miazzi MM, D'Agostino N, di Rienzo V et al (2020) Marginal grapevine germplasm from Apulia (Southern Italy) represents an unexplored source of genetic diversity. Agronomy 10:563. https://doi.org/10.3390/agronomy10040563

Migicovsky Z, Sawler J, Money D et al (2016) Genomic ancestry estimation quantifies use of wild species in grape breeding. BMC Genomics 17:478. https://doi.org/10.1186/s12864-016-2834-8

Migliaro D, De Lorenzis G, Di Lorenzo GS et al (2019) Grapevine non-vinifera genetic diversity assessed by simple sequence repeat markers as a starting point for new rootstock breeding programs. Am J Enol Vitic 70:390–397. https://doi.org/10.5344/ajev.2019.18054

Moralejo E, Borràs D, Gomila M et al (2019) Insights into the epidemiology of pierce's disease in vineyards of mallorca, Spain. Plant Pathol 68:1458–1471. https://doi.org/10.1111/ppa.13076

Myles S, Chia J-M, Hurwitz B et al (2010) Rapid genomic characterization of the genus vitis. PLoS One 5:e8219. https://doi.org/10.1371/journal.pone.0008219

OIV (2022) 2nd edition of the OIV Descriptor list for grape varieties and Vitis species. O.I.V., Paris. https://www.oiv.int/sites/default/files/2022-12/Code%202e%20edition%20Finale.pdf. Accessed 06 February 2025

Peakall R, Smouse PE (2012) GenALEx 6.5: genetic analysis in excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Péros J-P, Launay A, Peyrière A et al (2023) Species relationships within the genus Vitis based on molecular and morphological data. PLoS One 18:e0283324. https://doi.org/10.1371/journal.pone.0283324

Porras-Hurtado L, Ruiz Y, Santos C et al (2013) An overview of STRUCTURE: applications, parameter settings, and supporting software. Front Genet 4:1–13. https://doi.org/10.3389/fgene.2013.00098

Possamai T, Migliaro D, Gardiman M et al (2020) Rpv mediated defense responses in grapevine offspring resistant to Plasmopara viticola. Plants 9:781. https://doi.org/10.3390/plants9060781

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959. https://doi.org/10.1093/genetics/155.2.945

Rex F, Fechter I, Hausmann L, Töpfer R (2014) QTL mapping of black rot (Guignardia bidwellii) resistance in the grapevine rootstock 'Börner' (V. riparia Gm183 × V. cinerea Arnold). Theor Appl Genet 127:1667–1677. https://doi.org/10.1007/s00122-014-2329-4

Riaz S, Krivanek AF, Xu K, Walker MA (2006) Refined mapping of the pierce's disease resistance locus, PdR1, and sex on an extended genetic map of vitis rupestris × V. arizonica. Theor Appl Genet 113:1317–1329. https://doi.org/10.1007/s00122-006-0385-0

Riaz S, Tenscher AC, Graziani R et al (2009) Using marker-assisted selection to breed pierce's disease-resistant grapes. Am J Enol Vitic 60:199–207. https://doi.org/10.5344/ajev.2009.60.2.199

Riaz S, De Lorenzis G, Velasco D et al (2018a) Genetic diversity analysis of cultivated and wild grapevine (Vitis vinifera L.) accessions around the mediterranean basin and central Asia. BMC Plant Biol 18:137. https://doi.org/10.1186/s12870-018-1351-0

Riaz S, Huerta-Acosta K, Tenscher AC, Walker MA (2018b) Genetic characterization of vitis germplasm collected from the Southwestern US and Mexico to expedite pierce's disease-resistance breeding. Theor Appl Genet 131:1589–1602. https://doi.org/10.1007/s00122-018-3100-z

Riaz S, Tenscher AC, Heinitz CC et al (2020) Genetic analysis reveals an east-west divide within North American vitis species that mirrors their resistance to pierce's disease. PLoS One 15:7–10. https://doi.org/10.1371/journal.pone.0243445

Riaz S, Tenscher A, Walker MA (2023) Genetic mapping of pierce's disease resistance in germplasm collected from the Southwestern United States and Mexico. Am J Enol Vitic 74:0740026. https://doi.org/10.5344/ajev.2023.23006

Rodriguez MZ, Comin CH, Casanova D et al (2019) Clustering algorithms: a comparative approach. PLoS One 14:e0210236. https://doi.org/10.1371/journal.pone.0210236

Rogers JS (1972) Measures of genetic similarity and genetic distance. In: Studies in Genetics, VII. University of Texas Publication 7213, Austin, TX, pp 145–153

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425. https://doi.org/10.1093/oxfordjournals.molbev.a04045493683

Salem KFM, Sallam A (2015) Analysis of population structure and genetic diversity of Egyptian and exotic rice (Oryza sativa L.) genotypes. Comptes Rendus Biol 339:1–9. https://doi.org/10.1016/j.crvi.2015.11.003

Schwander F, Eibach R, Fechter I et al (2012) Rpv10: a new locus from the Asian vitis gene pool for pyramiding downy mildew resistance loci in grapevine. Theor Appl Genet 124:163–176. https://doi.org/10.1007/s00122-011-1695-4

Sefc KM, Regner F, Turetschek E et al (1999) Identification of microsatellite sequences in vitis riparia and their applicability for genotyping of different vitis species. Genome 42:367–373. https://doi.org/10.1139/g98-168

Sharma S, Upadhyaya HD, Varshney RK, Gowda CLL (2013) Pre-breeding for diversification of primary gene pool and genetic enhancement of grain legumes. Front Plant Sci 4. https://doi.org/10.3389/fpls.2013.00309

Singh N, Choudhury DR, Singh AK et al (2013) Comparison of SSR and SNP markers in Estimation of genetic diversity and population structure of Indian rice varieties. PLoS One 8:e84136. https://doi.org/10.1371/journal.pone.0084136

Sukumaran S, Rebetzke G, Mackay I et al (2022) Pre-breeding strategies. In: Reynolds MP, Braun H-J (eds) Wheat improvement. Springer International Publishing, Cham, pp 451–469. https://doi.org/10.1007/978-3-030-90673-3_25

Sun R, Lin F, Huang P, Zheng Y (2016) Moderate genetic diversity and genetic differentiation in the relict tree liquidambar formosana hance revealed by genic simple sequence repeat markers. Front Plant Sci 7. https://doi.org/10.3389/fpls.2016.01411

Tao L-L, Ting Y-J, Chen H-R et al (2023) Core collection construction of tea plant germplasm in Anhui Province based on genetic diversity analysis using simple sequence repeat markers. J Integr Agric 22:2719–2728. https://doi.org/10.1016/j.jia.2023.07.020

Tessier C, David J, This P et al (1999) Optimization of the choice of molecular markers for varietal identification in vitis vinifera L.

Theor Appl Genet 98:171–177. https://doi.org/10.1007/s001220051054

This P, Jung A, Boccacci P et al (2004) Development of a standard set of microsatellite reference alleles for identification of grape cultivars. Theor Appl Genet 109:1448–1458. https://doi.org/10.1007/s00122-004-1760-3

Thomas MR, Scott NS (1993) Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequence-tagged sites (STSs). Theor Appl Genet 86:985–990. https://doi.org/10.1007/BF00211051

Tian M, Li W, Luo P et al (2024) Genetic diversity analysis and core germplasm bank construction in cold resistant germplasm of rubber trees (Hevea brasiliensis). Sci Rep 14:14533. https://doi.org/10.1038/s41598-024-65464-9

Upadhyay A, Aher LB, Shinde MP et al (2013) Microsatellite analysis to rationalize grape germplasm in India and development of a molecular database. Plant Genet Resour 11:225–233. https://doi.org/10.1017/S1479262113000117

Vähä JP, Erkinaro J, Niemelä E, Primmer CR (2007) Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. Mol Ecol 16:2638–2654. https://doi.org/10.1111/j.1365-294X.2007.03329.x

van Heerden CJ, Burger P, Vermeulen A, Prins R (2014) Detection of downy and powdery mildew resistance QTL in a 'Regent' × 'RedGlobe' population. Euphytica 200:281–295. https://doi.org/10.1007/s10681-014-1167-4

Venuti S, Copetti D, Foria S et al (2013) Historical introgression of the downy mildew resistance gene Rpv12 from the Asian species vitis amurensis into grapevine varieties. PLoS ONE 8. https://doi.org/10.1371/journal.pone.0061228

Vezzulli S, Vecchione A, Stefanini M, Zulini L (2018) Downy mildew resistance evaluation in 28 grapevine hybrids promising for breeding programs in Trentino region (Italy). Eur J Plant Pathol 150:485–495. https://doi.org/10.1007/s10658-017-1298-2

Vezzulli S, Dolzani C, Migliaro D et al (2019) The Fondazione Edmund Mach grapevine breeding program for downy and powdery mildew resistances: toward a green viticulture. Acta Hortic 109–114. https://doi.org/10.17660/ActaHortic.2019.1248.16

Vezzulli S, Gramaje D, Tello J et al (2022) Genomic designing for biotic stress resistant grapevine. In: Genomic designing for biotic stress resistant fruit crops. Springer International Publishing, Cham, pp 87–255. https://doi.org/10.1007/978-3-030-91802-6_4

Viard F, Franck P, Dubois M-P et al (1998) Variation of microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species. J Mol Evol 47:42–51. https://doi.org/10.1007/PL00006361

VIVC (2024) Table of Loci for Traits in Grapevine Relevant for Breeding and Genetics. Accessed 5 Aug 2024. https://www.vivc.de/docs/dataonbreeding/20220920_Table%20of%20Loci%20for%20Traits%20in%20Grapevine.pdf

Wan Y, Schwaninger HR, Baldo AM et al (2013) A phylogenetic analysis of the grape genus (Vitis L.) reveals broad reticulation and concurrent diversification during neogene and quaternary climate change. BMC Evol Biol 13:141. https://doi.org/10.1186/1471-2148-13-141

Wang Y, Wu X, Li Y et al (2021) Identification and validation of a core single-nucleotide polymorphism marker set for genetic diversity assessment, fingerprinting identification, and core collection development in bottle gourd. Front Plant Sci 12. https://doi.org/10.3389/fpls.2021.747940

Wang P, Su J, Wu H et al (2023) Analysis of germplasm genetic diversity and construction of a core collection in Camellia oleifera C.Abel by integrating novel simple sequence repeat markers. Genet Resour Crop Evol 70:1517–1530. https://doi.org/10.1007/s10722-022-01519-y

Wapshere AJ, Helm KF (1987) Phylloxera and vitis: an experimentally testable coevolutionary hypothesis. Am J Enol Vitic 38:216–222. https://doi.org/10.5344/ajev.1987.38.3.216

Wilkerson D, Zou C, Sun Q et al (2025) Comparative genomics of Rpv3, a multiallelic downy mildew resistance locus in grapevine (Vitis sp). OENO One 59. https://doi.org/10.20870/oeno-one.2025.59.1.8244

Xue H, Yu X, Fu P et al (2021) Construction of the core collection of Catalpa fargesii f. duclouxii (Huangxinzimu) based on molecular markers and phenotypic traits. Forests 12:1518. https://doi.org/10.3390/f12111518

Zdunić G, Lukšić K, Nagy ZA et al (2020) Genetic structure and relationships among wild and cultivated grapevines from central Europe and part of the Western Balkan Peninsula. Genes (Basel) 11:1–15. https://doi.org/10.3390/genes11090962

Zecca G, Labra M, Grassi F (2020) Untangling the evolution of American wild grapes: admixed species and how to find them. Front Plant Sci 10. https://doi.org/10.3389/fpls.2019.01814

Zendler D, Schneider P, Töpfer R, Zyprian E (2017) Fine mapping of Ren3 reveals two loci mediating hypersensitive response against erysiphe necator in grapevine. Euphytica 213. https://doi.org/10.1007/s10681-017-1857-9

Zhang J, Hausmann L, Eibach R et al (2009) A framework map from grapevine V3125 (Vitis vinifera 'Schiava grossa' × 'Riesling') × rootstock cultivar 'börner' (Vitis riparia × Vitis cinerea) to localize genetic determinants of phylloxera root resistance. Theor Appl Genet 119:1039–1051. https://doi.org/10.1007/s00122-009-1107-1

Zini E, Dolzani C, Stefanini M et al (2019) R-Loci arrangement versus downy and powdery mildew resistance level: a vitis hybrid survey. Int J Mol Sci 20. https://doi.org/10.3390/ijms20143526

Žulj Mihaljević M, Maletić E, Preiner D et al (2020) Genetic diversity, population structure, and parentage analysis of Croatian grapevine germplasm. Genes (Basel) 11:737. https://doi.org/10.3390/genes11070737