Genetic structure and molecular diversity of Brazilian grapevine germplasm: management and use in breeding programs

Geovani Luciano de Oliveira[1], Anete Pereira de Souza[2,3], Fernanda Ancelmo de Oliveira[2], Maria Imaculada Zucchi[4], Lívia Moura de Souza[2], Mara Fernandes Moura[1]*

[1] Advanced Fruit Research Center, Agronomic Institute (IAC), Jundiaí, SP, Brazil

[2] Molecular Biology and Genetic Engineering Center (CBMEG), University of Campinas (UNICAMP), Campinas, SP, Brazil

[3] Department of Plant Biology, Biology Institute, University of Campinas (UNICAMP) UNICAMP, Campinas, SP, Brazil

[4] Laboratory of Conservation Genetics and Genomics, Agribusiness Technological Development of São Paulo (APTA), Piracicaba, SP, Brazil

* Corresponding author

E-mail: mouram@iac.sp.gov.br (MFM)

Abstract

The management of germplasm banks is complex, especially when many accessions are involved. Microsatellite markers are an efficient tool for assessing the genetic diversity of germplasm collections, optimizing their use in breeding programs. This study genetically characterizes a large collection of 420 grapevine accessions maintained at the Agronomic Institute of Campinas (IAC) (Brazil). The accessions were genotyped with 17 highly polymorphic microsatellite markers. Genetic data were analyzed to determine the genetic structure of the germplasm, quantify its allelic diversity, suggest the composition of a core collection and discover cases of synonymy, duplication and misnaming. A total of 304 alleles were obtained, and 334 unique genotypes were identified. The molecular profiles of 145 accessions were confirmed according to the literature and databases, and the molecular profiles of more than 100 genotypes were reported for the first time. The analysis of the genetic structure revealed different levels of stratification. The primary division was between accessions related to *Vitis vinifera* and *V. labrusca*, followed by their separation from wild grapevine. A core collection of 120 genotypes captured 100% of all detected alleles. The accessions selected for the core collection may be used in future phenotyping efforts, in genome association studies, and for conservation purposes. Genetic divergence among accessions has practical applications in grape breeding programs, as the choice of relatively divergent parents will maximize the frequency of progeny with superior characteristics. Together, our results can enhance the management of grapevine germplasm and guide the efficient exploitation of genetic diversity to facilitate the development of new grape cultivars for fresh fruits, wine and rootstock.

Introduction

Grapevine (*Vitis* spp.) is considered to be a major fruit crop globally based on hectares cultivated and economic value [1]. Grapevines are exotic species in Brazil but have become increasingly important in national fruit agriculture in recent years, transitioning from exclusive cultivation in temperate zones to a great alternative in tropical regions.

European grapevine, or *V. vinifera*, cultivars stand out in terms of their economic importance, being the most commonly planted worldwide and characterized by having excellent-quality fruits with wide morphological and genetic diversity. They are widely used for the production of fresh fruits, dried fruits, and juice and in the global fine and sparkling wine industry [2].

In Brazil, the American *V. labrusca* varieties and hybrids (*V. labrusca* x *V. vinifera*) thrive because of their vegetative characteristics, which are best adapted to the country's environmental conditions, with generally high humidity. In addition, due to their relatively high robustness, they are resistant to many diseases that affect grapevine in the country, resulting in production of relatively high volume, although of low quality, and have become dominant on Brazilian plantations [3,4].

The wild species of the genus *Vitis* have contributed evolutionarily through interspecific crossings, accidental or planned, to the adaptation of grapevine to the highly different conditions that its expansion has demanded. Hybrid varieties are characterized by greater resistance to pests and diseases than *V. vinifera* and by producing fruits with better organoleptic characteristics than American grapes. Crosses and natural mutations have greatly benefited from the possibility of vegetative propagation among grapevines, enabling the exploitation of different characteristics

over time, with noticeable variations in berries, flowers and leaves, further increasing the number of cultivars planted [2,5].

The starting point of any breeding program of a species is genetic variability, whether spontaneous or created. The manipulation of this variability with suitable methods leads to the safe obtainment of superior genotypes in relation to agronomic characteristics of interest [6]. Germplasm banks have a fundamental role in preserving this genetic variability but require the maintenance of accessions [7]. The quantification of the magnitude of genetic variability and its distribution between and within the groups of accessions that constitute germplasm banks is essential to promote its rational use and adequate management [8].

Most germplasm is derived from seeds, but for highly heterozygous plants, such as grapevines, this method is not suitable, with conservation most commonly occurring through the use of *ex situ* field collections. The germplasm banks involved in breeding programs are fundamental to the development of new materials. These collections generally have a large number of accessions, but only a small proportion of these resources are used in practice. The management of such collections becomes complex when many accessions are involved. Redundancy should be reduced to a minimum, the use of “true-to-type” plant material must be ensured, and the introduction of new accessions should be optimized [9]. Therefore, it is essential to identify and correct errors related to synonyms, homonyms and mislabeling that can occur during the introduction and propagation of plant material [10,11]. The genetic characterization of available genetic resources may permit the optimization of the use of these resources by grouping a sufficient number of accessions in a core collection to maximize the genetic diversity described in the whole collection [12].

Information on the genetic diversity available in germplasm banks is valuable for use in breeding programs because such information assists in the detection of combinations of accessions capable of producing progenies with maximum variability in characteristics of interest, guiding hybridization schemes [13].

The identification of grapevine cultivars has traditionally been based on ampelography, which is the analysis and comparison of the morphological characteristics of leaves, branches, shoots, bunches, and berries [14], but as this process is carried out on adult plants, a long period is necessary before accession identification can be completed. Since many synonyms or homonyms exist for cultivars [2], passport data are not always sufficient to certify identities, mainly in terms of the distinction of closely related cultivars, and errors can arise. Thus, the use of molecular markers has become an effective strategy for this purpose due to the high information content detected directly at the DNA level without environmental influence and in the early stages of plant development, allowing for faster and more accurate cultivar identification [15].

Microsatellites, or simple sequence repeats (SSRs), are among the most appropriate and efficient markers for genetic structure and conservation studies [16]. SSRs are highly polymorphic and transferable among several species of the genus *Vitis* [17]. Since SSRs provide unique fingerprints for cultivar identification [18], they have been used for genetic resource characterization [19,20], parentage analysis [21,22], genetic mapping [23,24], detection of quantitative trait loci (QTLs) [25] and assisted selection [26].

Because SSRs are highly reproducible and stable, they have allowed the development of several reference banks with grapevine variety genetic profiles from around the world. Access to these reference banks allows the exchange of information

between different research groups, significantly increasing international efforts related to the correct identification of grapevine genetic resources [27].

Considering the importance of viticulture and winemaking in Brazil, the Agronomic Institute of Campinas (IAC) has a *Vitis* spp. germplasm bank including wild *Vitis* species, interspecific hybrids and varieties of the main cultivated species (*V. vinifera*, *V. labrusca*, *V. bourquina*, *V. rotundifolia*) and varieties developed by the IAC.

Our objective in the present study was to describe the diversity and genetic structure of the *Vitis* spp. available in this germplasm bank using microsatellite markers. The accessions were characterized, and their molecular profiles were compared with the use of different literature and online databases. Here, we quantify the genetic diversity of this Brazilian germplasm and describe its genetic structure, and we suggest the composition of a core collection that would capture the maximum genetic diversity with a minimal sample size. We discuss perspectives related to the use of this information in germplasm management and conservation.

Materials and Methods

Plant material

A total of 420 accessions from the *Vitis* spp. Germplasm Bank of the IAC in Jundiaí, São Paulo (SP), Brazil, were analyzed (S1 Table). This germplasm encompasses more than ten species of *Vitis*, including commercial and noncommercial varieties of wine, table and rootstock grapes. Each accession consisted of three plants cultivated in an espalier system and pruned in August every year, leaving one or two buds per branch. Detailed data on the accessions are available in S1 Table.

DNA extraction

Total genomic DNA was extracted from young leaves homogenized in a TissueLyser (Qiagen, Valencia, CA, USA) following the cetyltrimethylammonium bromide (CTAB) method previously described by Doyle (1991) [28]. The quality and concentration of the extracted DNA were assessed using 1% agarose gel electrophoresis with comparison to known quantities of standard λ phage DNA (Invitrogen, Carlsbad, CA, USA).

Microsatellite analysis

A set of 17 grapevine SSR markers well characterized in previous studies were used, including ten developed by Merdinoglu et al. (2005) [29] (VVIn74, VVIr09, VVIp25b, VVIn56, VVIn52, VVIq57, VVIp31, VVIp77, VVIv36, VVIr21) and seven suggested by the guidelines of the European scientific community for universal grapevine identification, characterization, standardization, and exchange of information [30][31]: VVS2 [32], VVMD5, VVMD7 [33], VVMD25, VVMD27 [34], VrZAG62 and VrZAG79 [35]. One primer in each primer pair was 5' labeled with one of the following fluorescent dyes: 6-FAM, PET, NED or VIC.

Polymerase chain reaction (PCR) was performed using a three-primer labeling system [36] in a final volume of 10 μl containing 20 ng of template DNA, 0.2 μM of each primer, 0.2 mM of each dNTP, 2 mM MgCl2, 1× PCR buffer (20 mM Tris HCl [pH 8.4] and 50 mM KCl) and 1 U of Taq DNA polymerase. PCR amplifications were carried out with the following steps: 5 min of initial denaturation at 95°C followed by 35 cycles of 45 s at 94°C, 45 s at 56°C or 50°C (VVS2, VVMD7, VrZAG62 and VrZAG79), 1 min 30 s at 72°C, and a final extension step of 7 min at 72°C. Amplifications were checked with 3% agarose gels stained with ethidium bromide. The

amplicons were denatured with formamide and analyzed with an ABI 3500 (Applied Biosystems, Foster City, CA, USA) automated sequencer. The alleles were scored against the internal GeneScan-600 (LIZ) Size Standard Kit (Applied Biosystems, Foster City, CA, USA) using Geneious software v. 8.1.9 [37].

Genetic diversity analyses

Descriptive statistics for the genotyping data were generated using GenAlEx v. 6.5 [38] to indicate the number of alleles per locus (Na), effective number of alleles (Ne), observed heterozygosity (H_O), expected heterozygosity (H_E) and fixation index (F). GenAlEx software was also used to identify private (Pa) and rare alleles (frequency < 0.05). The polymorphism information content (PIC) and null allele frequency (*r*) were calculated for each locus using Cervus 3.0.7 [39]. Discriminating power (D*j*) [40] values were estimated to compare the efficiencies of microsatellite markers in varietal identification and differentiation.

Genetic structure analysis

To assess the genetic diversity among the genotypes, three different approaches were used: a Bayesian model-based approach, a distance-based model using a dissimilarity matrix and discriminant analysis of principal components (DAPC).

The model-based Bayesian analysis implemented in the software package STRUCTURE v. 2.3.4 [41] was used to determine the approximate number of genetic clusters (K) within the full dataset and to assign individuals to the most appropriate cluster. All simulations were performed using the admixture model, with 100,000 replicates for burn-in and 1,000,000 replicates for Markov chain Monte Carlo (MCMC)

processes in ten independent runs. The number of clusters (K) tested ranged from 1 to 10.

The online tool Structure Harvester [42] was used to analyze the STRUCTURE output, and the optimal K values were calculated using Evanno's ΔK *ad hoc* statistics [43]. The optimal alignment over the 10 runs for the optimal K values was obtained using the greedy algorithm in CLUMPP v.1.1.2 [44], and the results were visualized using DISTRUCT software v.1.1 [45].

Genetic distances between accessions were estimated on the basis of Rogers' genetic distance [46], and the resulting distance matrix was used to construct a dendrogram with the unweighted pair group method with arithmetic mean (UPGMA) algorithm, with 1,000 bootstrap replicates, implemented in the R package *poppr* [47]. The final dendrogram was formatted with iTOL v. 5.5 [48].

DAPC as implemented in the R package *adegenet* 2.1.2 [49,50] was also performed. The *find.clusters* function was used to detect the number of clusters in the germplasm, which runs successive K-means clustering with increasing numbers of clusters (K). We used 20 as the maximum number of clusters. The optimal number of clusters was estimated using the Bayesian information criterion (BIC), which reaches a minimum value when the best-supported assignment of individuals to the appropriate number of clusters is approached. DAPC results are presented as multidimensional scaling plots.

Accession name validation

To verify the trueness to type and identify misnamed genotypes, the molecular profiles obtained in this study were compared with the data contained in the following online databases: Vitis International Variety Catalogue (VIVC, www.vivc.de), Italian

Vitis Database (http://www.vitisdb.it), “Pl@ntGrape, le catalogue des vignes cultivées en France” (http://plantgrape.plantnet-project.org/fr) and the U.S. National Plant Germplasm System (NPGS, https://npgsweb.ars-grin.gov/gringlobal/search.aspx). For this comparison, the molecular profile of seven microsatellite loci (VVS2, VVMD5, VVMD7, VVMD25, VVMD27, VrZAG62, VrZAG79) adopted by the databases was used.

The allele sizes were first standardized for consistency with various references [51]. If an accession was not listed in these databases, it was verified in other scientific papers.

Core collection sampling

The R package *corehunter* 3.0 [52] was used to generate the core collection to represent the maximum germplasm genetic variability in a reduced number of accessions. Different core subsets were generated by changing the *size* parameter of the desired core collections. The sizes ranged from 0.1 to 0.3 for all datasets. For each core collection, the genetic diversity parameters were determined with GenAlEx v. 6.5 [38].

Ethics statement

We confirm that no specific permits were required to collect the leaves used in this study. This work was a collaborative study performed by researchers from the IAC (SP, Brazil), São Paulo's Agency for Agribusiness Technology (APTA, SP, Brazil) and the State University of Campinas (UNICAMP, SP, Brazil). Additionally, we confirm that this study did not involve endangered or protected species.

Results

Genetic diversity

Four hundred and twenty grapevine accessions of *Vitis* spp. were genotyped at 17 SSR loci (S1 Table). Of this total, ten accessions were excluded from the analysis due to amplification problems or polyploidy.

Among the remaining 410 accessions, we detected a total of 304 alleles at the 17 SSR loci analyzed (Table 1). The number of alleles per SSR locus (Na) ranged from 10 (VVIq57) to 24 (VVIp31), with an average of 17.88. The number of effective alleles per locus (Ne) varied from 2.39 (VVIq57) to 11.40 (VVIp31), with a mean value of 7.02.

Table 1. Genetic parameters of the 17 microsatellite loci obtained from 410 grapevine accessions.

Locus	**Na**	**Ne**	H_O	H_E	**PIC**	**D*j***	***r***
VVIn74	18	5.32	0.65	0.81	0.79	0.81	0.10
VVIr09	21	8.33	0.83	0.88	0.87	0.88	0.02
VVIp25b	21	4.15	0.48	0.75	0.73	0.76	0.21
VVIn56	12	3.27	0.60	0.69	0.65	0.69	0.06
VVIn52	13	7.87	0.58	0.87	0.86	0.87	0.20
VVIq57	10	2.39	0.56	0.58	0.52	0.58	0.00
VVIp31	24	11.14	0.88	0.91	0.90	0.91	0.01
VVIp77	23	8.22	0.76	0.87	0.86	0.88	0.06
VVIv36	15	4.74	0.79	0.78	0.76	0.79	0.00
VVIr21	17	6.74	0.83	0.85	0.83	0.85	0.01
VVS2	20	8.25	0.86	0.87	0.86	0.88	0.00
VVMD5	18	8.80	0.76	0.88	0.87	0.88	0.07
VVMD7	17	9.18	0.87	0.89	0.88	0.89	0.00
VVMD25	19	5.91	0.75	0.83	0.81	0.83	0.04
VVMD27	22	7.99	0.87	0.87	0.86	0.87	0.00
VrZAG62	18	8.32	0.84	0.88	0.86	0.88	0.01
VrZAG79	16	8.74	0.85	0.88	0.87	0.88	0.01

Total	304	119.42				
Mean	17.88	7.02	0.75	0.83	0.81	0.83

Number of alleles (Na), number of effective alleles (Ne), observed heterozygosity (H_O), expected heterozygosity (H_E), polymorphic information content (PIC), discrimination power (D*j*), estimated frequency of null alleles (*r*).

Across all of the accessions, the mean observed heterozygosity (H_O) was 0.75 (ranging from 0.48 to 0.88). The expected heterozygosity (H_E) was higher than the observed heterozygosity (H_O) for most loci, except for VVIv36. Among these loci ($H_O<H_E$), the probability of null alleles (*r*) was significantly high (>0.20) only for VVIp25b and VVIn52. The analysis revealed a high H_E level, ranging from 0.58 (VVIq57) to 0.91 (VVIp31), with a mean of 0.83.

The PIC estimates varied from 0.52 (VVIq57) to 0.90 (VVIp31), with a mean value of 0.81. The discrimination power (D*j*) was greater than 0.80 for 13 of the 17 loci, with the highest value for the VVIp31 locus (0.91). The D*j* values were high for 76.5% of the SSR markers used (>0.80). When the PIC and D*j* of each locus were analyzed together, 12 loci presented the highest values for both indexes (>0.80). In this study, the largest amount of information was provided by VVIp31, for which 24 alleles were detected showing a PIC and a D*j* ≥ 0.90.

Evaluation of genetic relationships and germplasm structure

The STRUCTURE analysis indicated the relatedness among the 410 accessions, with the highest ΔK value for K = 3, suggesting that three genetic clusters were sufficient to interpret our data (Fig 1A and 1B).

Fig 1. Genetic structure of the *Vitis* germplasm accessions obtained on the basis of 17 microsatellite markers. (A) The most probable number of groups (K) estimated based on the method described by Evanno et al. (2005) [43]. Bar graphs of the estimated membership proportions (Q) for each of the 410 accessions. Each accession is represented by a single vertical line, which is partitioned into colored segments in proportion to the estimated membership in each cluster. (B) Inferred genetic structure for K = 3. C1: genetic predominance of the species *V. vinifera*; C2: genetic predominance of the species *V. labrusca*; C3: genetic predominance of wild *Vitis* species; (C) inferred genetic structure for K = 7. VW: wine grape accessions related to *V. vinifera*; VT: table grape accessions related to *V. vinifera*; HI: IAC hybrids of table grapes; SS: accessions related to the Seibel series; VL: accessions related to the species *V. labrusca*; WV: accessions related to wild *Vitis* species; VR: *V. rotundifolia* accessions.

Based on a membership probability threshold of 70%, 207 accessions were assigned to cluster 1, 54 accessions were assigned to cluster 2, and 51 accessions were assigned to cluster 3. The remaining 98 accessions were assigned to the admixed group. The level of clustering (K = 3) is related to the main accession species. Cluster 1 was formed by accessions with the greatest relation to *V. vinifera*. Cluster 2 contained the accessions most related to *V. labrusca*. Accessions linked to wild *Vitis* species were allocated to cluster 3. All accessions assigned to the admixed group were identified as interspecific hybrids.

Of the 304 observed alleles, 227 were shared among the groups; the remaining 77 represented private alleles (Pa) in different groups of accessions (Fig 2). The VVMD27 locus had the largest number of private alleles of the 17 SSR markers used in

this study (9). Clusters 1 and 2, constituted by accessions related to the most cultivated species of grapevine, *V. vinifera* and *V. labrusca*, respectively, had the lowest number of private alleles (5 and 1, respectively). The greatest number of private alleles was found in cluster 3 (53), constituting 72.60% of the total private alleles.

Fig 2. Private allele (Pa) frequencies obtained from the genotyping of 410 grapevine accessions on the basis of 17 microsatellite loci. X-axis: Private allele frequencies; Y-axis: groups identified by STRUCTURE analyses at K = 3. The dashed line indicates the cutoff for the occurrence of rare alleles (frequency = 0.05).

The Evanno ΔK statistics indicated that other subdivisions (K = 4 and K = 7) could be pertinent. For K = 7 (Fig 1C), using a membership threshold of >70%, the accessions most related to *V. vinifera* were divided into two groups, one formed mainly by wine grapes (VW) (n = 91) and the other by table grapes (VT) (n = 33). In the VL group, *V. labrusca* hybrids were found (n = 36). The HI group was formed mostly by hybrids developed by the IAC (n = 50). The SS group (n = 51) was constituted by accessions of the Seibel series and hybrids that include cultivars of this complex in their genealogy. The accessions related to wild species used as rootstocks were assigned to the WV (n = 26) and VR (n = 11) groups, with the VR group including only *V. rotundifolia* accessions. The remaining 112 accessions were assigned to the admixed group and identified as interspecific hybrids. The results are also presented for K = 4 (S1 Fig.), for which no interpretable genetic groups were identified.

Additionally, the DAPC was performed with no prior information about the groupings of the evaluated accessions. Inspection of the BIC values (S2 Fig.) revealed that the division of the accessions into nine clusters was the most likely scheme to

explain the variance in this set of accessions. The allocation of individuals into clusters according to the DAPC showed several similarities to those achieved by STRUCTURE with K = 7, and both analyses showed the same pattern of clustering. Essentially, clusters 1 (magenta), 2 (yellow), 4 (brown), 8 (dark green) and 9 (green) of the DAPC reflect the gene pools SS (magenta), VT (yellow), HI (brown), VR (dark green) and WV (green) detected by STRUCTURE at K = 7, respectively. The groups VW and VL formed by STRUCTURE were subdivided in the DAPC analysis. The VW group (red) corresponded to DAPC clusters 5 (red) and 3 (orange), and the VL group (blue) reflected DAPC clusters 6 (blue) and 7 (purple).

Initially, the DAPC scatterplot showed the formation of three main distinct groups, with great genetic differentiation of clusters 8 (dark green) and 9 (green) from the others (Fig 3A). In a subsequent DAPC, outlier clusters 8 and 9 were removed to improve the visualization of the relationship of the other clusters (Fig 3B). In this second scatterplot, clusters 1 (magenta) and 7 (purple) showed greater genetic differentiation, with low variance within the groups, as well as no case of overlap with another cluster, indicating strong genetic structure.

Fig 3. DAPC scatterplots based on the K-means algorithm used to identify the proper number of clusters. Dots represent individuals, and the clusters are presented in different colors. The accessions were allocated into nine clusters: 1 (magenta), related to the Seibel series; 2 (yellow), related to table grape accessions of *V. vinifera*; 3 (orange) and 5 (red), related to wine grape accessions of *V. vinifera*; 4 (brown), predominance of IAC hybrids; 6 (blue) and 7 (purple), related to the species *V. labrusca*; 8 (dark green), related to wild *Vitis* species; and 9 (green), *V. rotundifolia* accessions. (A) DAPC with all samples included. (B) DAPC excluding clusters 8 and 9.

Finally, we constructed a dendrogram using the UPGMA method from the distance matrix based on Rogers' distance to confirm the relationships among the accessions (Fig 4). The dendrogram showed a pattern that was consistent with those from the two analyses described above. The groups formed by the accessions of wild species were separated. The wild species *V. rotundifolia* was the most distantly related to other *Vitis* species, as seen in the DAPC. There was also a strong separation between accessions related to *V. labrusca* and other accessions. The wine grape accessions of *V. vinifera* were mainly concentrated at the top of the dendrogram, while the table grape accessions of this species were found at the bottom. However, the other hybrids (IAC, Seibel series and others) were scattered among all the groups formed by the dendrogram except for groups of wild species.

Fig 4. UPGMA dendrogram based on Rogers' distance calculated from the dataset of 17 microsatellite markers across 410 grapevine accessions. Accessions colored according to species group.

Validation analysis of molecular profiles

The identification of one hundred and forty-five accessions was validated through matches with the data available in the literature and databases. The results also confirmed matches to reference profiles of clones based on somatic mutations. Another 42 accessions showed molecular profiles that matched a validated reference profile of a different prime name, indicating mislabeling (S1 Table).

The molecular profiles of the remaining 223 accessions did not match any available reference profile. This accession group included wild species and cultivars

from grapevine breeding programs in Brazil (the IAC and Embrapa), the United States, and France (Seibel series). The molecular profiles of more than 100 hybrids developed by the IAC were reported for the first time.

The accessions ‘101-14’, ‘Bailey’, ‘Black July’, ‘Carlos’, ‘Carman’, ‘Castelão’, ‘Catawba Rosa’, ‘Elvira’, ‘Moscatel de Alexandria’ and ‘Regent’ showed a different profile than the reference profile of the same name and did not match any other available reference profile. However, more morphological and source information is needed to validate their identification. To avoid possible confusion, these accessions were indicated as “Unknown”.

After correcting the mislabeling, 22 cases of duplicates were identified, all with accessions of the same name and same molecular profile. Accessions identified with different names but having the same molecular profile were classified as synonyms. Thirty-one synonymous groups were elucidated in this study (S2 Table). Some accessions classified as “Unknown” showed genetic profiles identical to accessions that did not match any available reference profile; examples can be seen in synonymous groups 1, 2, 5 and 6 in S2 Table.

Core collection

Three different core collections were constructed with a size ranging from 10 to 30% of the entire dataset to identify the smallest set of accessions that would be able to represent as much of the available genetic diversity as possible (Table 2). Core 3, composed of 120 accessions, managed to capture 100% of the 304 detected alleles, while the smallest core collection (0.1) managed to capture 243 alleles, approximately 20% less than the total number of alleles detected. The genetic diversity index values obtained for the core collection were similar to or higher than those for the entire

germplasm. The H_O values ranged from 0.64 (Core 1) to 0.70 (Core 3); the value for Core 3 was similar to that detected for all 410 accessions (0.75). The three nuclear collections showed H_E and Ne values higher than those observed for the entire dataset. The H_E values for all core collections were 0.85, while Ne ranged from 134.53 (Core 2) to 137.69 (Core 3). The values of Ne and H_E are related to allele frequencies, and low values of allele frequencies generate even lower values when squared. With a reduction in the number of accessions (N), the low-frequency alleles (allele frequency between 0.05 to 0.25) and rare alleles (frequency less than 0.05) showed an increase in frequency, resulting in an increase in Ne and H_E.

Table 2. SSR diversity within each core collection compared with that of the entire dataset (IAC collection).

	Size	**N**	**Na**	**Ne**	**H_O**	**H_E**	**Total SSR diversity captured (%)**
Core 1	0.1	41	243	136,223	0,639	0,858	79,93
Core 2	0.2	82	275	134,530	0,695	0,852	90,46
Core 3	0.3	120	304	137,696	0,700	0,855	100
IAC collection	1.0	410	304	119,421	0,755	0,833	100

Number of accessions (N), number of alleles (Na), number of effective alleles (Ne), observed heterozygosity (H_O), expected heterozygosity (H_E)

All clusters detected in the STRUCTURE analysis and DAPC are represented in Core 3. In particular, in the STRUCTURE analysis at K = 3, 49 accessions were in cluster 1, 12 were in cluster 2, 29 were in cluster 3 and 30 were in the admixture group, representing 41, 10, 24 and 25% of Core 3, respectively.

Discussion

Genetic diversity

The results of this study revealed high levels of genetic diversity among the evaluated accessions. The observed high genetic diversity was expected since the grape germplasm from the IAC includes varieties with very diverse origins, wild species, and different intra- and interspecific hybrids. These results are similar to those found in other Brazilian germplasm banks characterized by their abundance of hybrid varieties [53,54].

The high number of alleles per locus identified (~18) is likely due to the taxonomic amplitude of the germplasm since a relatively high number of low-frequency alleles were found in wild species accessions. Lamboy and Alpha (1998) [17], when analyzing the diversity of 110 accessions belonging to 21 species of *Vitis* and 4 hybrids, detected 24.4 alleles per locus, a greater quantity than that observed in this study, showing that taxonomically broader accessions contribute to a greater number of alleles.

Most loci had lower H_O values than those expected from the randomized union of gametes (H_E), except for VVIv36. For these loci, the probability of null alleles was positive but significantly high (> 0.20) for only VVIp25b and VVIn52. This suggests that at these loci, some of the apparent homozygotes could be heterozygous, with one allele being visible and the other not. Such null alleles can occur when mutations prevent the linking of primers to the target region [55].

The high number of alleles obtained by the 17 SSR primer set positively impacted the PIC and discrimination power (D*j*). According to the classification of Botstein et al. (1980) [56], all the loci used can be considered highly informative (PIC > 0.50).

Structure and genetic relationship of accessions

The genetic structure was mostly impacted by two factors that are difficult to separate: clear discrimination based on species and human usage as wine, table or rootstock grapes, as previously noted by Laucou et al. (2018) [57] and Emanuelli et al. (2013) [58]. A population structure analysis using the software STRUCTURE revealed the presence of three primary clusters in our set of accessions based on the species *V. vinifera* and *V. labrusca* and wild *Vitis*. However, a large number of accessions were not assigned and remained in a large admixed group, evidencing the genetic complexity of the analyzed plant material. Many of the accessions in the admixed group are interspecific selections developed by crossing wild American species resistant to phylloxera, downy mildew or powdery mildew with European grapevine varieties. The intra- and interspecific crossings carried out during breeding cycles in search of novelties and hybrid vigor promote the miscegenation of grapevine cultivars, resulting in hybrids with a heterogeneous genetic composition.

Our results demonstrated the greatest number of private alleles in cluster 3, which is composed of the wild germplasm (Fig 2). This finding confirms that the wild accessions are important reservoirs of genetic variation, with the potential for incorporating new materials into breeding programs in response to the demand for the development of cultivars with different characteristics. Wild grape germplasm is a potential source of unique alleles and provides the breeder with a set of genetic resources that may be useful in the development of cultivars resistant to pests and diseases, tolerant to abiotic stresses, and that even show enhanced productivity, which makes their conservation of paramount importance [59].

At K = 7, the STRUCTURE results agree with the groupings identified through DAPC (Fig 1C and Fig 3), in which the genotypes from *V. vinifera* were separated

according to their use. The VW group was composed mainly of wine grapes, such as the accessions ‘Syrah’, ‘Merlot Noir’, ‘Chenin Blanc’, ‘Petit Verdot’ and ‘Cabernet Sauvignon’, which showed associations with a membership greater than 95%, corresponding to DAPC clusters 5 (red) and 3 (orange). The *V. vinifera* accessions of table grapes of ‘Centennial Seedless’, ‘Aigezard’, ‘Moscatel de Hamburgo’, and ‘Italia’ and their mutations ‘Benitaka’, ‘Rubi’ and ‘Brazil’ were found in the VT group. This group corresponded to cluster 2 (yellow) in the DAPC. In the UPGMA dendrogram, the *V. vinifera* accessions were also completely separated in terms of use; the wine grapes were located at the top, and the table grapes were located at the bottom.

In the DAPC and UPGMA dendrograms, two groups were differentiated to a greater extent than the others (Fig 3 and Fig 4), with these groups being formed mainly of wild grapes often used as rootstocks. This likely occurred because few rootstocks used worldwide contain part of the *V. vinifera* genome [9], while practically all table and wine grape hybrids present in this germplasm contain a part of it. In both analyses, the *V. rotundifolia* accessions constituted the most divergent group. This species is the only one in the germplasm belonging to the *Muscadinia* subgenus, which contains plants with 2n = 40 chromosomes, while the others belong to the *Euvitis* subgenus, with 2n = 38 chromosomes. The high genetic divergence between *V. rotundifolia* and the species in the *Euvitis* subgenus was also observed by Costa et al. (2017) [60] through the use of RAPD molecular markers and by Miller et al. (2013) [61] through SNPs. The species *V. rotundifolia* is resistant to several grapevine pests and diseases [62] and is an important source of genetic material in the development of cultivars and rootstocks adapted to the most diverse environmental conditions and with tolerance and/or resistance to biotic and abiotic factors.

The HI and SS groups at K = 7 were formed by interspecific hybrids, and these groups corresponded to clusters 4 and 1 in the DAPC. The HI group mainly included table grape hybrids developed by the IAC, while the SS group contained accessions of the Seibel series and hybrids with varieties of this series in their genealogy. The Seibel series is in fact a generic term referring to several hybrid grapes developed in France at the end of the 19th century by Albert Seibel from crosses between European *V. vinifera* varieties and wild American *Vitis* species to develop phylloxera-resistant cultivars with characteristics of fine European grapes [63]. The combinations of alleles of different *Vitis* species certainly created unique genetic pools, which explains the genetic distinction of this group of accessions.

The *V. labrusca* hybrids formed distinct groups in the three analyses. In the STRUCTURE analysis, these accessions constituted cluster 2 at K = 3 and the VL group at K = 7. In the DAPC, this accession group was subdivided into two clusters (6 and 7) according to use (Fig 3). Cluster 6 contained only table grape cultivars, including ‘Eumelan’, ‘Niabell’, ‘Highland’, ‘Niagara’ and their mutations, while grape cultivars for processing, including ‘Isabella’, ‘Ives’ and ‘Concord Precoce’ were included in cluster 7.

The similarities in the STRUCTURE, DAPC and UPGMA dendrograms provide important knowledge for the management of germplasm diversity. The identification of divergent groups guides crossings in breeding programs, facilitating the appropriate combination of parents to obtain progeny with wide genetic variability, allowing the maximization of heterosis and making it possible to obtain individuals with superior characteristics. Information about the available genetic diversity is valuable because if properly explored, it can reduce vulnerability to genetic erosion through the avoidance

of crosses between genetically related genotypes while also accelerating genetic progress related to characteristics of importance to grape growth [64].

Identification analysis: misnamed and synonymous cases

Considering the vast diversity of names for the different varieties of grapevine, standardization is necessary. Errors due to homonyms, synonyms, differences in spelling and misnamed accessions impede the estimation of the real number of different accessions existing in grapevine collections, with a negative impact on grapevine breeding programs. Therefore, the verification of true-to-type accessions is indispensable [30]. For grapevine, a 6-SSR genotyping system has been established as a useful tool for identification and parentage analysis, allowing the allele length of varieties to be comparably scored by different institutions [31,51].

In this study, 42 cases of misnaming were found by comparing the molecular profiles obtained with the information available in the literature and databases (S1 Table). The molecular profile of the accession ‘Cabernet Franc’ corresponded to the cultivar Merlot Noir, and the molecular profile of the accession ‘Merlot Noir’ corresponded to the cultivar Cabernet Franc, clearly indicating an exchange of nomenclature between these accessions. ‘Cabernet Franc’ is one of the parents of ‘Merlot Noir’, and some morphological traits of these two cultivars are quite similar [65], which certainly contributed to the occurrence of this mistake.

The molecular profile of the accession ‘Magoon’ matched that of ‘Regale’ in the present study, and ‘Regale’ had a similar molecular profile to those obtained by Schuck et al. (2011) [66] and Riaz et al. (2008) [67], indicating that ‘Magoon’ was misnamed at the time of introduction and that both accessions were the cultivar Regale. In Brazil, the same case of misnaming was also reported by Schuck et al. (2011) [66]. A misspelling

case was observed for the accession ‘Pedro Ximenez’, corresponding to ‘Pedro Gimenez’; both cultivars are classified by the VIVC as wine grapes with white berries. However, despite the similar names and some comparable characteristics, the genealogies of these cultivars are completely different, being easily distinguished by microsatellite marker analysis due to the different molecular profiles generated.

The accessions ‘Armenia I70060’ and ‘Armenia I70061’ were labeled according to their country of origin, Armenia, during their introduction. Through microsatellite marker analysis, these accessions were identified as ‘Aigezard’ and ‘Parvana’, respectively. A similar situation was observed for the accession ‘Moscatel Suiça’, corresponding to ‘Muscat Bleu’ from Switzerland; this accession was likely also labeled according to its country of origin.

Additionally, 31 synonymous groups were identified (S2 Table). Cases of synonymy could correspond to clones of the same cultivar that show phenotypic differences due to the occurrence of somatic mutations [68,69]. Mobile elements are known to generate somatic variation in vegetatively propagated plants such as grapevines [70,71]. Carrier et al. (2012) [70] observed that insertion polymorphism caused by mobile elements is the major cause of mutational events related to clonal variation. In grape, retrotransposon-induced insertion into *VvmybA1*, a homolog of *VlmybA1-1*, is the molecular basis of the loss of pigmentation in a white grape cultivar of *V. vinifera* due to the lack of anthocyanin production [72].

The detection of somatic mutations is very difficult with a low number of microsatellite markers, especially when they are located in noncoding regions of the genome [73]. This was the case for synonymous groups 19, 20, 21, 22 in S2 Table, such as the cultivar Italia and its mutations ‘Rubi’, ‘Benitaka’ and ‘Brasil’, which differ in terms of the color of berries, with white, pink, red and black fruits, respectively, and are

cultivated as distinct cultivars in Brazil. This was also the case for 'Pinot Gris', a variant with gray berries arising from 'Pinot Noir', which has black berries. The mutations that occurred in the cultivar Niagara can be distinguished in terms of the color, size, and shape of the berries, and they may even lead to a lack of seeds, such as the apyrenic accession 'Niagara Seedless' or 'Rosinha' [74].

The accession 'Tinta Roriz' was identified as a synonym of the cultivar Tempranillo Tinto in this study; this synonym is already registered in the VIVC and is widely used in regions of Portugal [75]. The wild species *V. doaniana* and *V. berlandieri* have the same molecular profile, indicating a case of mislabeling; certainly, some mistake was made during the acquisition of these materials, and the same genotype was propagated with different names.

The occurrence of misidentification is common, especially for old clonal species such as *Vitis* spp., and it can occur during any stage of accession introduction and maintenance. It has been observed that 5 to 10% of the grape cultivars maintained in grape collections are incorrectly identified [76,77]. In a new place, a certain genotype may receive a new name, confusing samples and the maintenance of accessions in germplasm banks [78]. The correct identification of accessions is fundamental to optimize germplasm management and for the use of germplasm in ongoing breeding programs since related genotypes will not be chosen for field experiments or controlled crosses. The identification of the existence of synonyms, homonyms and misnamed accessions is essential to prevent future propagation and breeding errors [66] and in helping to reduce germplasm maintenance costs without the risk of losing valuable genetic resources. Since morphological descriptors are highly influenced by environmental factors, molecular analyses can support identification. SSR markers have often been considered very efficient at the cultivar level since they can be easily used to

distinguish different cultivars; however, they are less effective in differentiating clones [9]. The results of molecular analysis should not replace ampelographic observations but should be integrated with such observations, mainly for the identification of somatic mutations.

In this study, 223 accessions with molecular profiles did not match any available reference profile. The largest subset of accessions was from the Brazilian grapevine breeding program of the IAC, with 109 molecular profiles described for the first time. The identification and description of unreported molecular profiles is important for regional and national viticulture and ensures the institution's intellectual property rights over these cultivars. The information obtained in this study will contribute to international cooperation to correctly identify grape germplasm and will allow the inclusion of new molecular profiles of Brazilian grapevine cultivars in the database.

Development of a core collection

The intention for the development of a core collection is to represent the genetic diversity of the entire germplasm in a reduced set of accessions that is feasible to handle. The efficiency of the approach based on SSR profiles in identifying a core collection was already demonstrated for grapevine by Le Cunff et al. (2008) [79], Cipriani et al. (2010) [80], Emanuelli et al. (2013) [58] and Migliaro et al. (2019) [81].

In this study, 120 accessions were necessary to capture all the allelic diversity of the whole collection, which is equivalent to approximately 30% of all accessions (Table 2). In *V. vinifera* subsp. *sativa* core collections, the same result was obtained with smaller percentages of individuals, from 4 to 15% [58,79,80]. According to Le Cunff et al. (2008) [79], the use of only cultivated genotypes of *V. vinifera* subsp. *sativa* is one of

the reasons for the small number of individuals in the nuclear collection since cultivated genotypes tend to be less diverse than wild counterparts [12,82].

Migliaro et al. (2019) [81] analyzed 379 grapevine rootstock accessions and managed to represent their full allelic richness with a core collection containing 30% of the accessions, a result similar to that observed in this study. According to these authors, the large number of individuals in the core collection can be related to the number of varieties belonging to different *Vitis* species and the high genetic variability detected. These are likely the same reasons for the need for a high number of genotypes in our core collection, since the *Vitis* spp. Germplasm Bank of the IAC also includes accessions belonging to different *Vitis* species and many interspecific hybrids that have complex pedigrees (derived from crosses among three or more species). The comparison of different methods used to form core collections is not easy, as the analyses are rarely performed in the same way, and the original collections rarely include the same global diversity of species [79].

Among the 120 genotypes in Core 3, 82 were identified as interspecific hybrids, with 13 of them being non-*vinifera* varieties. This large number of interspecific hybrids in the core collection can be explained by their predominance in germplasm; in addition, many of them have a complex pedigree, which certainly combines alleles of different species of *Vitis*. Regarding the other genotypes in Core 3, 31 were identified as *V. vinifera* cultivars and seven as wild *Vitis*.

The core collection was constructed to provide a logical subset of germplasm for examination when the entire collection cannot be used. Complementary criteria, such as phenotypic, agronomic and adaptive traits, should be associated with the core collection to make it more fully representative. Finally, this core collection will be useful for the

development of new breeding strategies, future phenotyping efforts and genome-wide association studies.

Conclusions

The set of microsatellite markers used in our study was efficient in assessing the genetic diversity present in the evaluated germplasm, showing that there is wide genetic diversity to be explored. Knowledge of allelic diversity and how it is structured in germplasm collections is essential for future breeding efforts, whether through traditional hybridization or new breeding technologies. Taken together, our results can be used to enhance the management of grapevine germplasms and provide molecular data from a large set of genetic resources that contribute to expanding existing database information.

Acknowledgments

The authors are grateful to the São Paulo Research Foundation (FAPESP) and Coordination for the Improvement of Higher Education Personnel (CAPES) for supporting the project and researchers.

References

1. Torregrosa L, Vialet S, Adivèze A, Iocco-Corena P, Thomas MR. Grapevine (Vitis vinifera L.). In: Wang K, editor. Agrobacterium Protocols Methods in Molecular Biology. New York, NY: Springer; 2015. pp. 177–194. doi:10.1007/978-1-4939-1658-0_15
2. This P, Lacombe T, Thomas MR. Historical origins and genetic diversity of wine

grapes. Trends Genet. 2006;22: 511–519. doi:10.1016/j.tig.2006.07.008

3. Camargo UA, Tonietto J, Hoffmann A. Progressos na Viticultura Brasileira. Rev Bras Frutic. 2011;33: 144–149. doi:10.1590/S0100-29452011005000028

4. Tecchio MA, Hernandes JL, Pires EJP, Terra MM, Moura MF. Cultivo da videira para mesa, vinho e suco. 2nd ed. In: Pio R, editor. Cultivo de fruteiras de clima temperado em regiões subtropicais e tropicais. 2nd ed. Lavras, BR: UFLA; 2018. pp. 512–585.

5. Khadivi A, Gismondi A, Canini A. Genetic characterization of Iranian grapes (Vitis vinifera L.) and their relationships with Italian ecotypes. Agrofor Syst. 2019;93: 435–447. doi:10.1007/s10457-017-0134-1

6. Alves J da Si, Ledo CA da S, Silva S de O e, Pereira VM, Silveira D de C. Divergência genética entre genótipos de bananeira no estado do Rio de Janeiro. Magistra. 2012;24: 116–122.

7. Nass LL, Sigrist MS, Ribeiro CS da C, Reifschneider FJB. Genetic resources: the basis for sustainable and competitive plant breeding. Crop Breed Appl Biotechnol. 2012;12: 75–86. doi:10.1590/s1984-70332012000500009

8. Manechini JRV, Costa JB da, Pereira BT, Carlini-Garcia LA, Xavier MA, Landell MG de A, et al. Unraveling the genetic structure of Brazilian commercial sugarcane cultivars through microsatellite markers. PLoS One. 2018;13. doi:10.1371/journal.pone.0195623

9. Laucou V, Lacombe T, Dechesne F, Siret R, Bruno JP, Dessup M, et al. High throughput analysis of grape genetic diversity as a tool for germplasm collection management. Theor Appl Genet. 2011;122: 1233–1245. doi:10.1007/s00122-010-1527-y

10. Leão PCS, Riaz S, Graziani R, Dangl GS, Motoike SY, Walker MA.

Characterization of a Brazilian grape germplasm collection using microsatellite markers. Am J Enol Vitic. 2009;60: 517–524.

11. Doulati-Baneh H, Mohammadi SA, Labra M. Genetic structure and diversity analysis in Vitis vinifera L. cultivars from Iran using SSR markers. Sci Hortic (Amsterdam). 2013;160: 29–36. doi:10.1016/j.scienta.2013.05.029

12. De Souza LM, Guen V Le, Cerqueira-Silva CBM, Silva CC, Mantello CC, Conson ARO, et al. 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. 2015;10: 1–20. doi:10.1371/journal.pone.0134607

13. Mohammadi SA, Prasanna BM. Analysis of genetic diversity in crop plants - Salient statistical tools and considerations. Crop Sci. 2003;43: 1235–1248. doi:10.2135/cropsci2003.1235

14. Boursiquot JM, This P. Les nouvelles techniques utilisées en ampélographie: informatique et marquage. J Int des Sci la Vigne du Vin. 1996;Special: 13–23.

15. Roychowdhury R, Taoutaou A, Hakeem KR, Ragab M, Gawwad A, Tah J. Molecular Marker-Assisted Technologies for Crop Improvement. 2014; 241–258. doi:10.13140/RG.2.1.2822.2560

16. Jarne P, Lagoda PJL. Microsatellites, from molecules to populations and back. Trends Ecol Evol. 1996;11: 424–429. doi:10.1016/0169-5347(96)10049-5

17. Lamboy WF, Alpha CG. Using simple sequence repeats (SSRs) for DNA fingerprinting germplasm accessions of grape (Vitis L.) species. Journal of the American Society for Horticultural Science. 1998. pp. 182–188. doi:10.21273/jashs.123.2.182

18. Cipriani G, Frazza G, Peterlunger E, Testolin R. Grapevine fingerprinting using

microsatellite repeats. Vitis. 1994;33: 211–215.

19. Buhner-Zaharieva T, Moussaoui S, Lorente M, Andreu J, Núñez R, Ortiz JM, et al. Preservation and molecular characterization of ancient varieties in spanish grapevine germplasm collections. Am J Enol Vitic. 2010;61: 557–562. doi:10.5344/ajev.2010.09129

20. Martín JP, Borrego J, Cabello F, Ortiz JM. Characterization of Spanish grapevine cultivar diversity using sequence-tagged microsatellite site markers. Genome. 2003;46: 10–18. doi:10.1139/g02-098

21. De Lorenzis G, Imazio S, Biagini B, Failla O, Scienza A. Pedigree reconstruction of the Italian grapevine aglianico (Vitis vinifera L.) from Campania. Mol Biotechnol. 2013;54: 634–642. doi:10.1007/s12033-012-9605-9

22. De Lorenzis G, Casas G Las, Brancadoro L, Scienza A. Genotyping of Sicilian grapevine germplasm resources (V. vinifera L.) and their relationships with Sangiovese. Sci Hortic (Amsterdam). 2014;169: 189–198. doi:10.1016/j.scienta.2014.02.028

23. Guo Y, Lin H, Liu Z, Zhao Y, Guo X, Li K. SSR and SRAP marker-based linkage map of Vitis vinifera L. Biotechnol Biotechnol Equip. 2014;28: 221–229. doi:10.1080/13102818.2014.907996

24. Grando MS, Bellin D, Edwards KJ, Pozzi C, Stefanini M, Velasco R. Molecular linkage maps of Vitis vinifera L. and Vitis riparia Mchx. Theor Appl Genet. 2003;106: 1213–1224. doi:10.1007/s00122-002-1170-3

25. Doligez A, Bouquet A, Danglot Y, Lahogue F, Riaz S, Meredith CP, et al. Genetic mapping of grapevine (Vitis vinifera L.) applied to the detection of QTLs for seedlessness and berry weight. Theor Appl Genet. 2002;105: 780–795. doi:10.1007/s00122-002-0951-z

26. Saifert L, Sánchez-Mora FD, Assumpção WT, Zanghelini JA, Giacometti R, Novak EI, et al. Marker-assisted pyramiding of resistance loci to grape downy mildew. Pesqui Agropecu Bras. 2018;53: 602–610. doi:10.1590/S0100-204X2018000500009

27. Lefort F, Roubelakis-Angelakis KA. The greek vitis database: A multimedia web-backed genetic database for Germplasm management of vitis resources in Greece. J Wine Res. 2000;11: 233–242. doi:10.1080/713684241

28. Doyle J. DNA Protocols for Plants. Mol Tech Taxon. 1991; 283–293. doi:10.1007/978-3-642-83962-7_18

29. Merdinoglu D, Butterlin G, Bevilacqua L, Chiquet V, Adam-Blondon AF, Decroocq S. Development and characterization of a large set of microsatellite markers in grapevine (Vitis vinifera L.) suitable for multiplex PCR. Mol Breed. 2005;15: 349–366. doi:10.1007/s11032-004-7651-0

30. This P, Dettweiler E. EU-project genres CT96 No81: European vitis database and results regarding the use of a common set of microsatellite markers. Acta Hortic. 2003;603: 59–66. doi:10.17660/ActaHortic.2003.603.3

31. International Organisation of Vine and Wine (OIV). 2nd edition of the OIV Descriptor list for grape varieties and Vitis species. Paris: O.I.V.; 2001.

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

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

34. Bowers JE, Dangl GS, Meredith CP. Development and characterization of

additional microsatellite DNA markers for grape. Am J Enol Vitic. 1999;50: 243□246.

35. Sefc KM, Regner F, Turetschek E, Glössl J, Steinkellner H. Identification of microsatellite sequences in Vitis riparia and their applicability for genotyping of different Vitis species. Genome. 1999;42: 367–373. doi:10.1139/g98-168

36. Schuelke M. An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol. 2000;18: 233–234. doi:10.1038/72708

37. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28: 1647–1649. doi:10.1093/bioinformatics/bts199

38. Peakall R, Smouse PE. GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics. 2012;28: 2537–2539. doi:10.1093/bioinformatics/bts460

39. Kalinowski ST, Taper ML, Marshall TC. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol. 2007;16: 1099–1106. doi:10.1111/j.1365-294X.2007.03089.x

40. Tessier C, David J, This P, Boursiquot JM, Charrier A. Optimization of the choice of molecular markers for varietal identification in Vitis vinifera L. Theor Appl Genet. 1999;98: 171–177. doi:10.1007/s001220051054

41. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155: 945–59.

42. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method.

Conserv Genet Resour. 2012;4: 359–361. doi:10.1007/s12686-011-9548-7

43. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol Ecol. 2005;14: 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x

44. Jakobsson M, Rosenberg NA. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007;23: 1801–1806. doi:10.1093/bioinformatics/btm233

45. Rosenberg NA. DISTRUCT: A program for the graphical display of population structure. Mol Ecol Notes. 2004;4: 137–138. doi:10.1046/j.1471-8286.2003.00566.x

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

47. Kamvar ZN, Tabima JF, Gr□unwald NJ. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ. 2014;2014: 1–14. doi:10.7717/peerj.281

48. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47: W256–W259. doi:10.1093/nar/gkz239

49. Jombart T. Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics. 2008;24: 1403–1405. doi:10.1093/bioinformatics/btn129

50. Jombart T, Ahmed I. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. Bioinformatics. 2011;27: 3070–3071.

doi:10.1093/bioinformatics/btr521

51. This P, Jung A, Boccacci P, Borrego J, Botta R, Costantini L, et al. Development of a standard set of microsatellite reference alleles for identification of grape cultivars. Theor Appl Genet. 2004;109: 1448–1458. doi:10.1007/s00122-004-1760-3

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

53. Schuck MR, Moreira FM, Guerra MP, Voltolini JA, Grando MS, Silva AL da. Molecular characterization of grapevine from Santa Catarina, Brazil, using microsatellite markers. Pesqui Agropecuária Bras. 2009;44: 487–495. doi:10.1590/s0100-204x2009000500008

54. Leão PC de S, Cruz CD, Motoike SY. Diversity and genetic relatedness among genotypes of Vitis spp. using microsatellite molecular markers. Rev Bras Frutic. 2013;35: 799–808. doi:10.1590/s0100-29452013000300017

55. Pollefeys P, Bousquet J. Molecular genetic diversity of the French – American grapevine hybrids cultivated in North America. Genome. 2003;46: 1037–1048. doi:10.1139/g03-076

56. Botstein D, White RL, Skolnick M, Davis RW. Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. Am J Hum Genet. 1980;32: 314–331.

57. Laucou V, Launay A, Bacilieri R, Lacombe T, Andre MT De, Hausmann L, et al. Extended diversity analysis of cultivated grapevine Vitis vinifera with 10K genome-wide SNPs. PLoS One. 2018;13: 1–27. doi:https://doi.org/10.1371/journal.pone.0192540

58. Emanuelli F, Lorenzi S, Grzeskowiak L, Catalano V, Stefanini M, Troggio M, et

al. Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. BMC Plant Biol. 2013;13: 1–17. doi:10.1186/1471-2229-13-39

59. Hajjar R, Hodgkin T. The use of wild relatives in crop improvement□: A survey of developments over the last 20 years The use of wild relatives in crop improvement□: A survey of developments over the last 20 years. Euphytica. 2007;156: 1–13. doi:10.1007/s10681-007-9363-0

60. Costa AF, Teodoro PE, Bhering LL, Tardin FD, Daher RF. Molecular analysis of genetic diversity among vine accessions using DNA markers. Genet Mol Res. 2017;16: 1–9. doi:10.4238/gmr16029586

61. Miller AJ, Matasci N, Schwaninger H, Aradhya MK, Prins B, Zhong G-Y, et al. Vitis Phylogenomics: Hybridization Intensities from a SNP Array Outperform Genotype Calls. Wang T, editor. PLoS One. 2013;8: 1–11. doi:10.1371/journal.pone.0078680

62. Kellow A V., McDonald G, Corrie AM, Heeswijck R. In vitro assessment of grapevine resistance to two populations of phylloxera from Australian vineyards. Aust J Grape Wine Res. 2002;8: 109–116. doi:10.1111/j.1755-0238.2002.tb00219.x

63. Pommer C V. Uva: Tecnologia de produção, pós-colheita, mercado. Porto Alegre: Cinco Continentes; 2003.

64. Leão PC de S, Motoike SY. Genetic diversity in table grapes based on RAPD and microsatellite markers. Pesqui Agropecu Bras. 2011;46: 1035–1044. doi:10.1590/S0100-204X2011000900010

65. Boursiquot JM, Lacombe T, Laucou V, Julliard S, Perrin FX, Lanier N, et al. Parentage of merlot and related winegrape cultivars of southwestern france:

Discovery of the missing link. Aust J Grape Wine Res. 2009;15: 144–155. doi:10.1111/j.1755-0238.2008.00041.x

66. Schuck MR, Biasi LA, Mariano AM, Lipski B, Riaz S, Walker MA. Obtaining interspecific hybrids, and molecular analysis by microsatellite markers in grapevine. Pesqui Agropecu Bras. 2011;46: 1480–1488. doi:10.1590/S0100-204X2011001100009

67. Riaz S, Tenscher AC, Smith BP, Ng DA, Walker MA. Use of SSR markers to assess identity, pedigree, and diversity of cultivated muscadine grapes. J Am Soc Hortic Sci. 2008;133: 559–568. doi:10.21273/jashs.133.4.559

68. Walker AR, Lee E, Robinson SP. Two new grape cultivars, bud sports of Cabernet Sauvignon bearing pale-coloured berries, are the result of deletion of two regulatory genes of the berry colour locus. Plant Mol Biol. 2006;62: 623–635. doi:10.1007/s11103-006-9043-9

69. Vignani R, Bowers JE, Meredith CP. Microsatellite DNA polymorphism analysis of clones of Vitis vinifera “Sangiovese.” Sci Hortic (Amsterdam). 1996;65: 163–169. doi:10.1016/0304-4238(95)00865-9

70. Carrier G, Le Cunff L, Dereeper A, Legrand D, Sabot F, Bouchez O, et al. Transposable elements are a major cause of somatic polymorphism in vitis vinifera L. PLoS One. 2012;7: 1–10. doi:10.1371/journal.pone.0032973

71. Fernandez L, Torregrosa L, Segura V, Bouquet A, Martinez-Zapater JM. Transposon-induced gene activation as a mechanism generating cluster shape somatic variation in grapevine. Plant J. 2010;61: 545–557. doi:10.1111/j.1365-313X.2009.04090.x

72. Kobayashi S, Goto-Yamamoto N, Hirochika H. Retrotransposon-Induced Mutations in Grape Skin Color. Science. 2004;304: 982.

doi:10.1126/science.1095011

73. Zulini L, Fabro E, Peterlunger E. Characterisation of the grapevine cultivar Picolit by means of morphological descriptors and molecular markers. Vitis. 2005;44: 35–38.

74. Souza JSI, Martins FP. Viticultura Brasileira: principais variedades e suas características. Piracicaba: FEALQ; 2002.

75. Ibáñez J, Muñoz-Organero G, Hasna Zinelabidine L, Teresa de Andrés M, Cabello F, Martínez-Zapater JM. Genetic origin of the grapevine cultivar tempranillo. Am J Enol Vitic. 2012;63: 549–553. doi:10.5344/ajev.2012.12012

76. De Andrés MT, Cabezas JA, Cervera MT, Borrego J, Martínez-Zapater JM, Jouve N. Molecular characterization of grapevine rootstocks maintained in germplasm collections. Am J Enol Vitic. 2007;58: 75–86.

77. Dettweiler E. The grapevine herbarium as an aid to grapevine identification- First results. Vitis. 1992;31: 117–120.

78. Moura EF, Farias Neto JT de, Sampaio JE, Silva DT da, Ramalho GF. Identification of duplicates of cassava accessions sampled on the North Region of Brazil using microsatellite markers. Acta Amaz. 2013;43: 461–467. doi:10.1590/s0044-59672013000400008

79. Le Cunff L, Fournier-Level A, Laucou V, Vezzulli S, Lacombe T, Adam-Blondon AF, et al. Construction of nested genetic core collections to optimize the exploitation of natural diversity in Vitis vinifera L. subsp. sativa. BMC Plant Biol. 2008;8. doi:10.1186/1471-2229-8-31

80. Cipriani G, Spadotto A, Jurman I, Gaspero G Di, Crespan M, Meneghetti S, et al. 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. 2010;121: 1569–1585. doi:10.1007/s00122-010-1411-9

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

82. Bartsch D, Lehnen M, Clegg J, Pohl-Orf M, Schuphan I, Ellstrand NC. Impact of gene flow from cultivated beet on genetic diversity of wild sea beet populations. Mol Ecol. 1999;8: 1733–1741. doi:10.1046/j.1365-294x.1999.00769.x

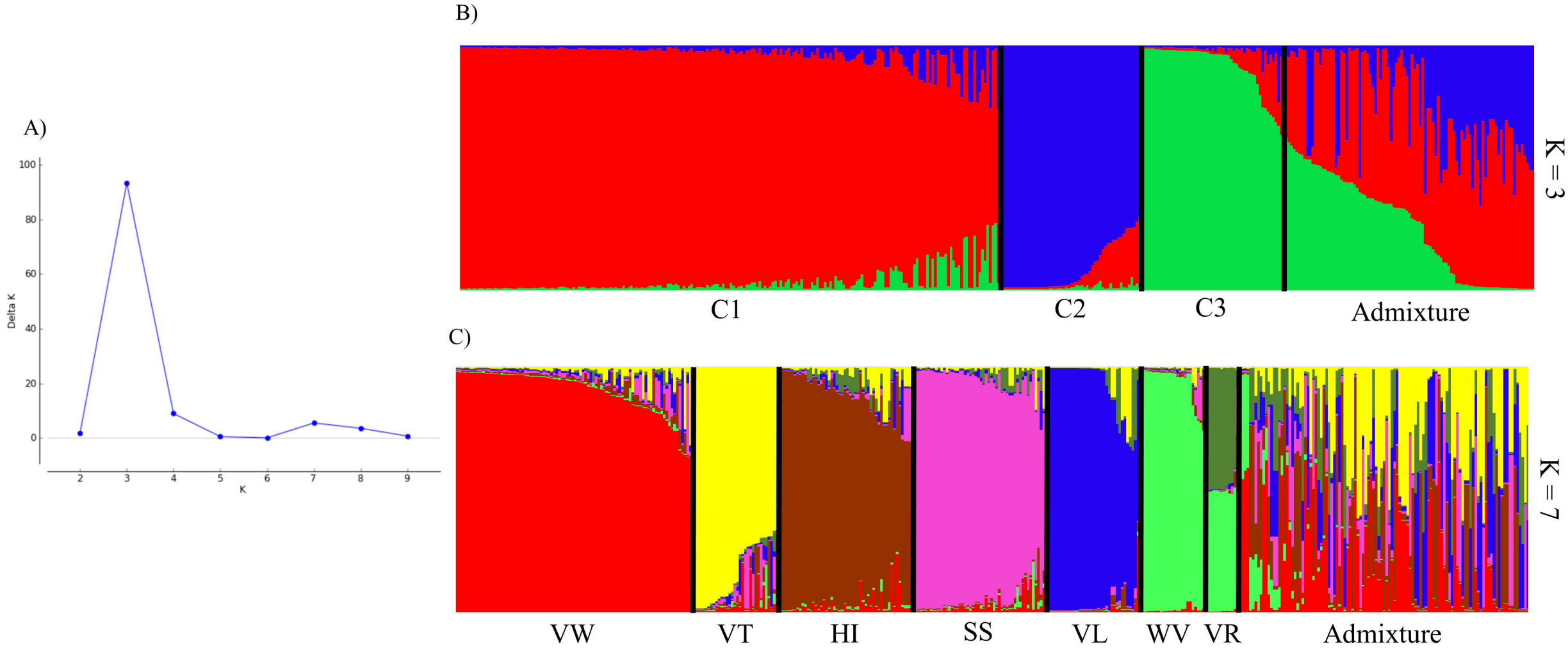
Supporting Information

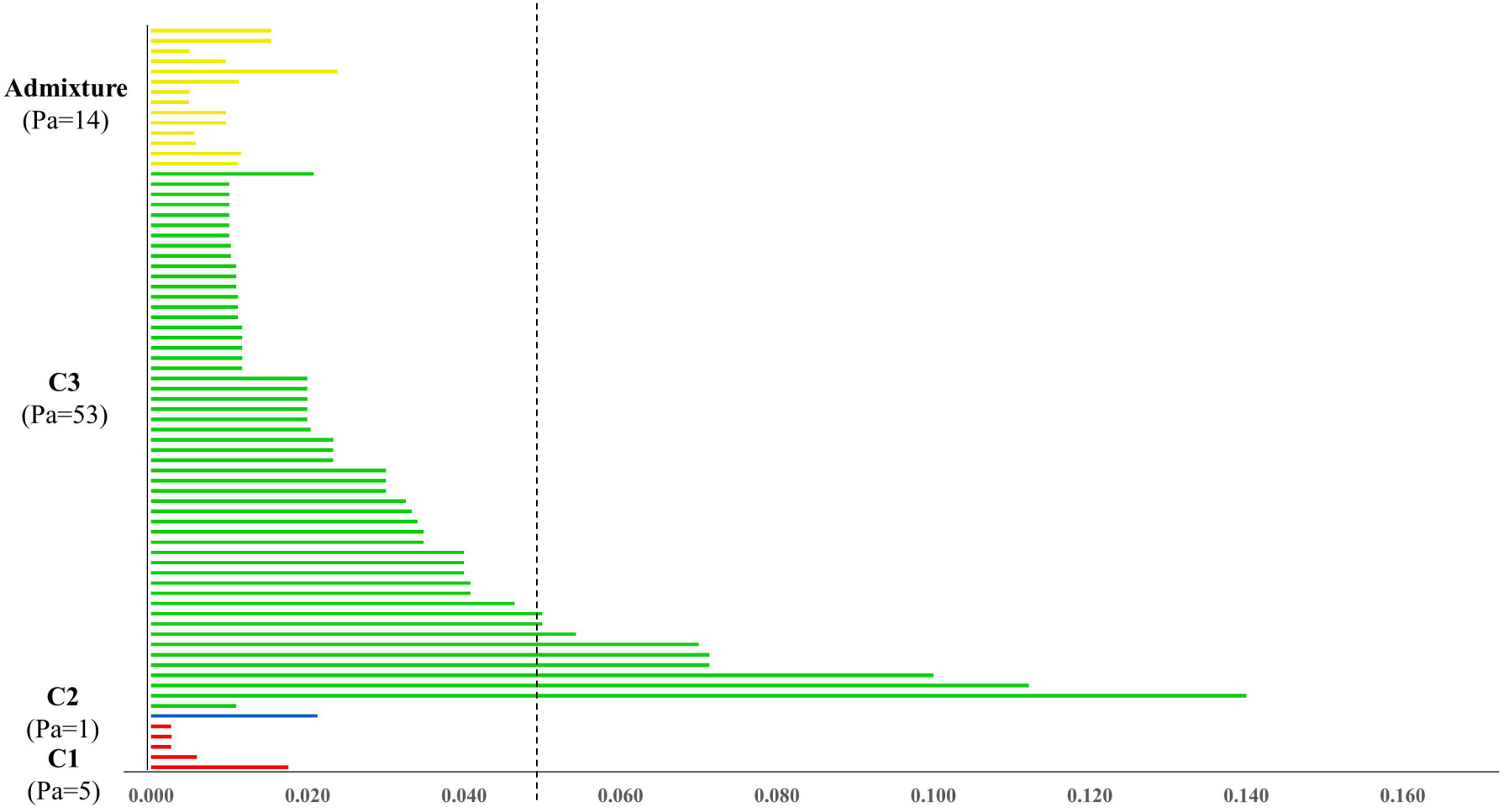
S1 Table. Genotypes used in this study. Detailed characteristics: SSR matches, usage, country of origin, species, group assignments according to the DAPC and STRUCTURE analysis at K = 3, K = 4 and K = 7, core collection composition and genotyping results obtained with the 17 microsatellite markers.

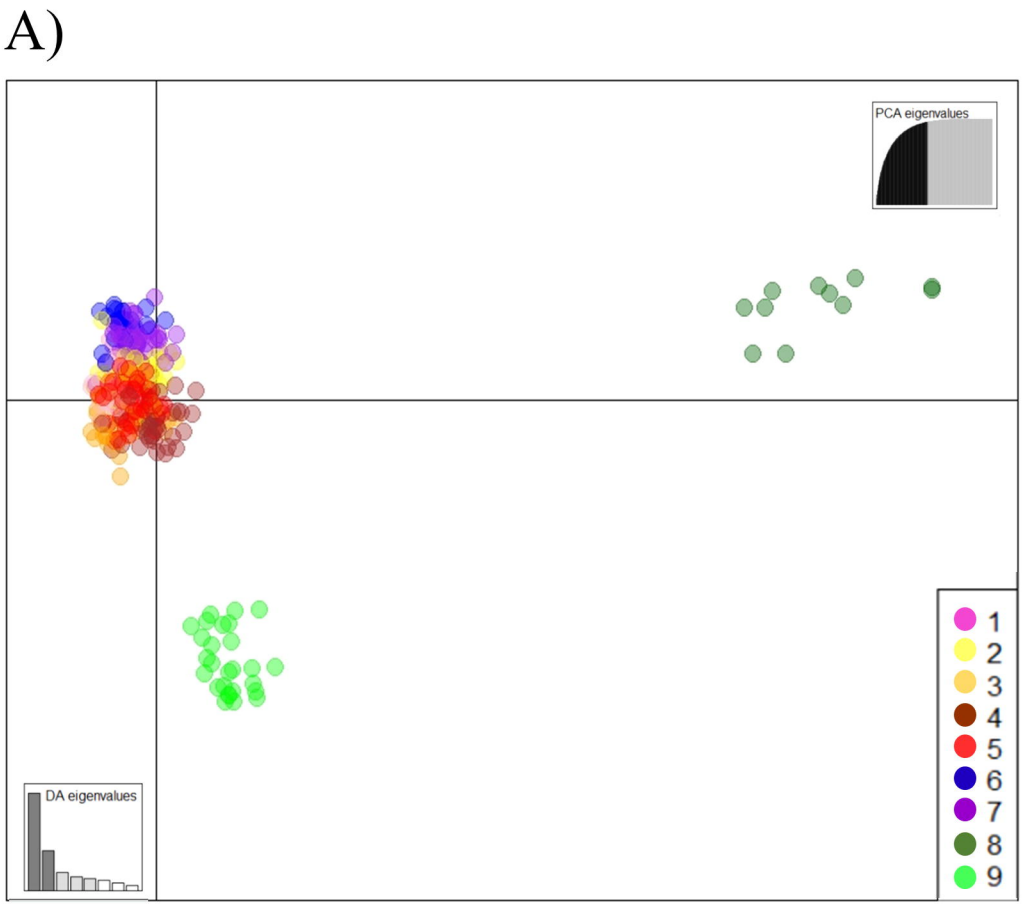
S2 Table. List of synonyms found in the *Vitis* spp. Germplasm Bank of the Agronomic Institute of Campinas (IAC) by SSR analysis.

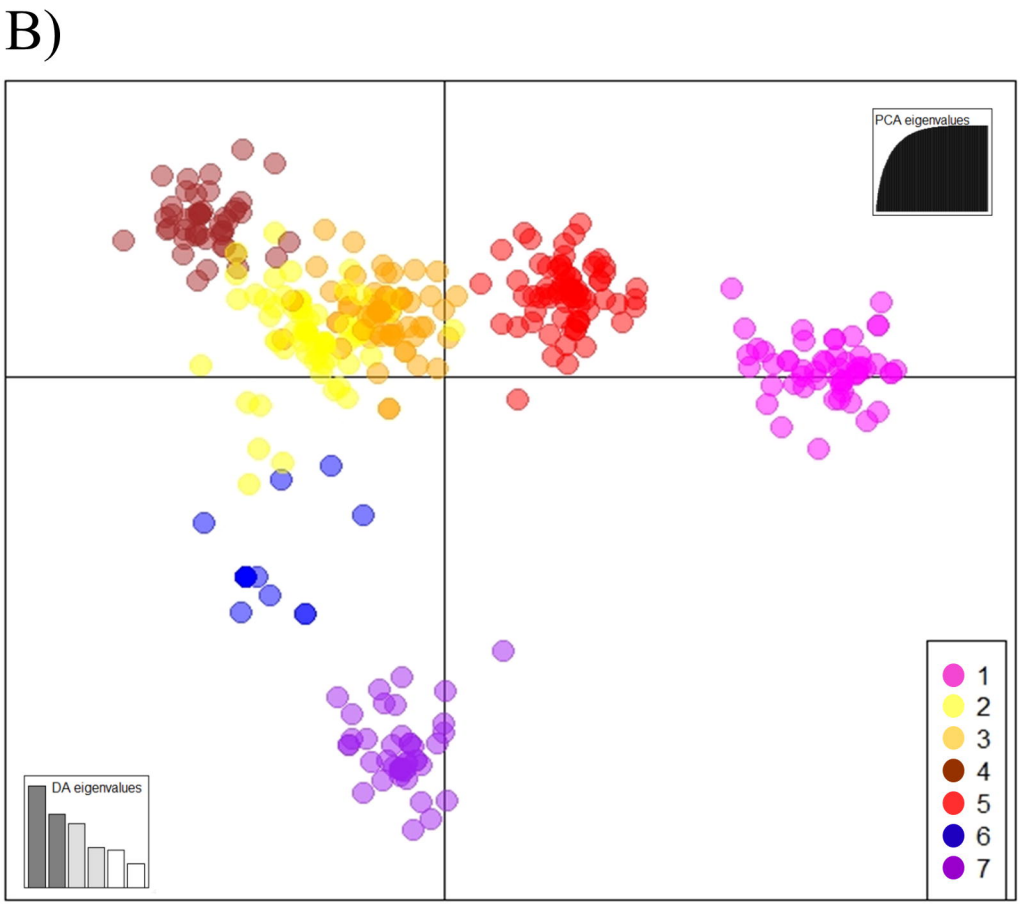
S1 Fig. Bayesian clustering results. Bar graph of the estimated membership coefficients (Q) according to STRUCTURE software for K = 4 for each of the 410 grapevine accessions evaluated.

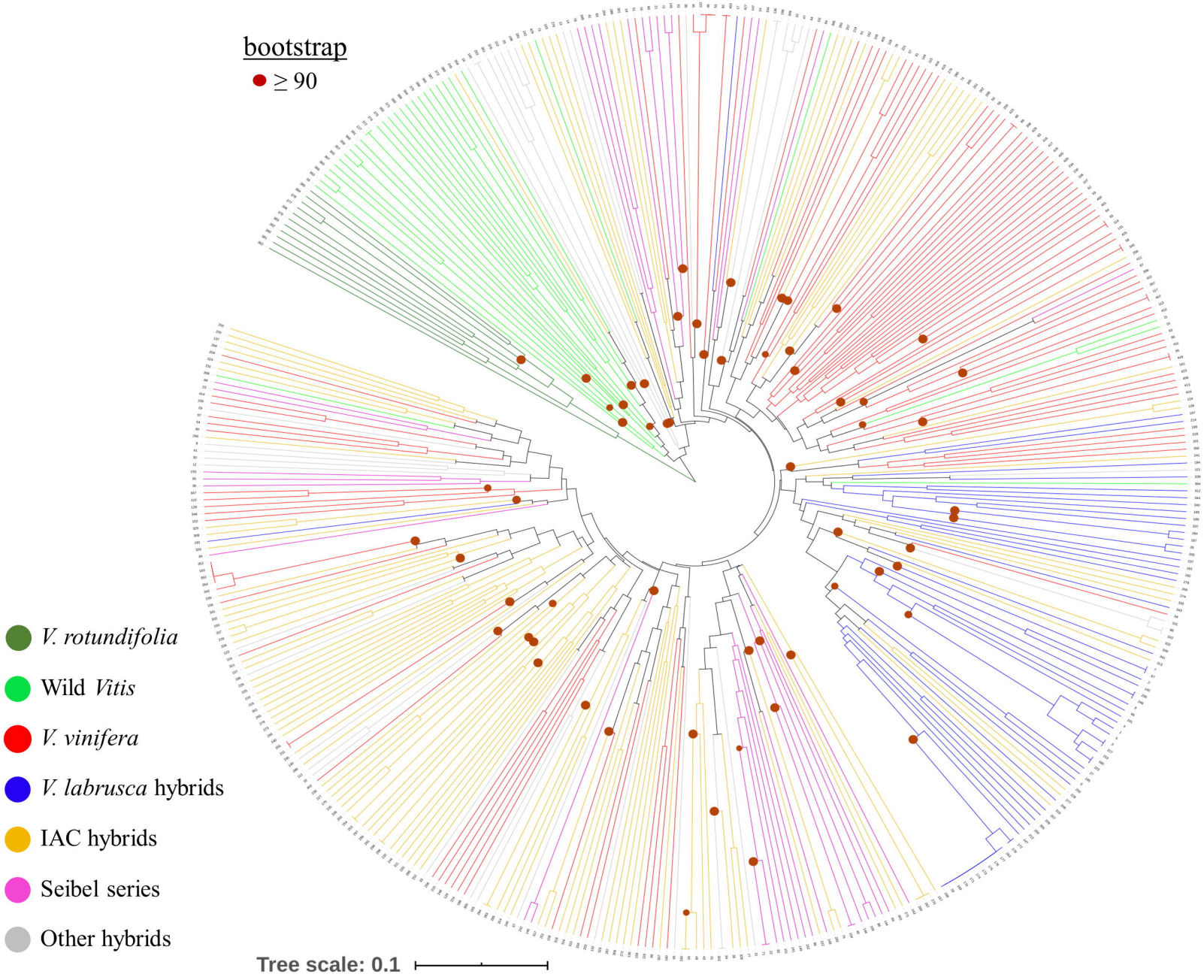
S2 Fig. Bayesian information criterion (BIC) values for different numbers of clusters. The accepted true number of clusters was nine.


A)
Delta K
100
80
60
40
20
0
2
3
4
5
6
7
8
9
K
B)
C1
C2
C3
Admixture
K = 3
C)
VW
VT
HI
SS
VL
WV
VR
Admixture
K = 7


Admixture
(Pa=14)
C3
(Pa=53)
C2
(Pa=1)
C1
(Pa=5)
0.000
0.020
0.040
0.060
0.080
0.100
0.120
0.140
0.160


A)
PCA eigenvalues
DA eigenvalues
1
2
3
4
5
6
7
8
9


B)
PCA eigenvalues
DA eigenvalues
1
2
3
4
5
6
7


bootstrap
≥ 90
V. rotundifolia
Wild Vitis
V. vinifera
V. labrusca hybrids
IAC hybrids
Seibel series
Other hybrids
Tree scale: 0.1