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 410 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 fruits of excellent quality 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 410 accessions from the *Vitis* spp. Germplasm Bank of the IAC in Jundiaí, São Paulo (SP), Brazil, were analyzed. 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 clonally propagated plants, sustained in an espalier system and pruned in August every year, leaving one or two buds per branch. For sampling, were collected young leaves of a single plant from each accession. 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 [22,29–31] were used, including ten developed by Merdinoglu et al. (2005) [32] (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 [33,34]: VVS2 [35], VVMD5, VVMD7 [36], VVMD25, VVMD27 [37], VrZAG62, and VrZAG79 [38]. One primer in each primer pair was 5' labeled with one of the following fluorescent dyes: 6-FAM, PET, NED, or VIC. Additional information about the loci is available in S2 Table.

Polymerase chain reaction (PCR) was performed using a three-primer labeling system [39] 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 using 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 [40].

Genetic diversity analyses

Descriptive statistics for the genotyping data were generated using GenAlEx v. 6.5 [41] 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), discriminating power (D*j*), and null allele frequency (*r*) were calculated to evaluate the efficiency and discriminatory potential of each microsatellite marker. Polymorphism information content (PIC) was calculated using Cervus 3.0.7 [42] according to the expression $PIC = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n} \sum_{j=i+1}^{n} 2p_i^2 p_j^2$, where n is the number of alleles, and p_i and p_j are the frequencies of the i^{th} and j^{th} alleles [43]. Discriminating power (D*j*) values were estimated to compare the efficiencies of microsatellite markers in varietal identification and differentiation. This parameter was calculated in accordance with the formula as follows: $D_j = 1 - C_j = 1 - \sum_{i=1}^{I} p_i \frac{Np_i - 1}{N - 1}$, where D_j is the probability that two randomly selected samples have different and distinct banding patterns, p_i is the frequency of the i^{th} pattern revealed by each marker, N is the number of samples analyzed, and I is the total number of patterns generated by each marker [44].

The null allele frequency (*r*) was estimated using Cervus 3.0.7. By definition, a microsatellite null allele is any allele at a microsatellite locus that consistently fails to amplify to detectable levels via the polymerase chain reaction (PCR) [45]. Cervus 3.0.7 uses a iterative likelihood approach [46], in which the presence of null allele homozygotes is not taken into consideration initially but is added in later optimization rounds. This method avoids overestimating the frequency of a null allele if samples fail to amplify for reasons other than the presence of nulls [45].

Genetic structure analysis

To assess the overall germplasm structuring, three approaches with different grouping criteria that do not require *a priori* assignment of individuals to groups 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 [47] was used to determine the approximate number of genetic clusters (K) within the full dataset and to assign individuals to the most appropriate cluster. STRUCTURE can identify subsets of individuals by detecting allele frequency differences within the data by assigning individuals to sub-populations based on analysis of likelihoods. The process begins by randomly assigning individuals to a pre-determined number of groups, after which variant frequencies are estimated in each group and individuals re-assigned based on those frequency estimates. This process is repeated many times in the burn-in process that results in a progressive convergence toward reliable allele frequency estimates in each population and membership probabilities of individuals to a population. During each analysis, membership coefficients summing to one are assigned to individuals for each group. If admixture is

considered, membership coefficients are generated across multiple clusters. The assumptions are that loci are unlinked and populations are in Hardy-Weinberg Equilibrium (HWE) [48]. Additionally, a "hierarchical STRUCTURE analysis" [49] was applied in this study by running STRUCTURE subsequently for each identified cluster separately to reveal any underlying structure, as suggested by Pritchard et al. (2007) [50].

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 [51] was used to analyze the STRUCTURE output, and the optimal K values were calculated using Evanno's ΔK *ad hoc* statistics [52]. The optimal alignment over the 10 runs for the optimal K values was obtained using the greedy algorithm in CLUMPP v.1.1.2 [53], and the results were visualized using DISTRUCT software v.1.1 [54]. Based on the posterior probability of membership (q), we classified individuals who showed $q \geq 0.70$ as members of a given cluster. In contrast, accessions with a membership of $q < 0.70$ were classified as admixed. This procedure was performed to avoid individuals constrained to belong to any of the given number (K) of clusters.

Distance-based methods proceed by calculating a pairwise distance matrix, the entries of which provide the distance between every pair of individuals. This matrix may then be represented using some convenient graphical representation, such as a dendrogram, and clusters may be identified by eye [47]. Genetic distances between accessions were estimated on the basis of Rogers' genetic distance [55], and the resulting distance matrix was used to construct a dendrogram with the neighbor-joining

algorithm [56], with 1,000 bootstrap replicates implemented in the R package *poppr* [57]. The principle of this method is to find pairs of operational taxonomic units that minimize the total branch length at each stage of clustering starting with a star-like tree [56]. The final dendrogram was formatted with iTOL v. 5.5 [58].

DAPC as implemented in the R package *adegenet* 2.1.2 [59,60] was also performed. DAPC is a multivariate analysis that does not rely on the assumption of HWE, the absence of linkage disequilibrium, or specific models of molecular evolution to identify clusters within genetic data. In DAPC, data are first transformed using a principal components analysis (PCA), after which a discriminant analysis (DA) is performed for the retained principal components. This process ensures that variables submitted to DA are perfectly uncorrelated and that their number is less than that of the analyzed individuals [61]. 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 [62]. 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 [63] was used to generate the core collection to represent the maximum germplasm genetic variability in a reduced number of accessions. Different samples were generated by changing the *size* parameter of the desired core collections to identify the subset of genotypes that could capture the entire diversity of alleles. The sizes ranged from 0.1 to 0.3 for all datasets. For each sample, the genetic diversity parameters were determined with GenAlEx v. 6.5 [41].

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 ten grapevine accessions of *Vitis* spp. were analyzed at 17 SSR loci (S1 Table), and a total of 304 alleles were detected (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	
SE*	0.93	0.57	0.03	0.02	0.02	0.02	

298 Number of alleles (Na), number of effective alleles (Ne), observed heterozygosity (H_O),
299 expected heterozygosity (H_E), polymorphic information content (PIC), discrimination
300 power (D*j*), estimated frequency of null alleles (*r*).
301 *Standard error of mean values.
302
303 Across all the accessions, the mean observed heterozygosity (H_O) was 0.75
304 (ranging from 0.48 to 0.88). The expected heterozygosity (H_E) was higher than the
305 observed heterozygosity (H_O) for most loci, except for VVIv36. Among these loci
306 ($H_O<H_E$), the probability of null alleles (*r*) was significantly high (>0.20) only for
307 VVIp25b and VVIn52. The analysis revealed a high H_E level, ranging from 0.58
308 (VVIq57) to 0.91 (VVIp31), with a mean of 0.83.
309 The PIC estimates varied from 0.52 (VVIq57) to 0.90 (VVIp31), with a mean
310 value of 0.81. The discrimination power (D*j*) was greater than 0.80 for 13 of the 17 loci,
311 with the highest value for the VVIp31 locus (0.91). The D*j* values were high for 76.5%
312 of the SSR markers used (>0.80). When the PIC and D*j* of each locus were analyzed
313 together, 12 loci presented the highest values for both indexes (>0.80). In this study, the
314 largest amount of information was provided by VVIp31, for which 24 alleles were
315 detected showing a PIC and a D$j \geq 0.90$.
316

317 Evaluation of genetic relationships and germplasm structure

318
319 The STRUCTURE analysis indicated the relatedness among the 410 accessions,
320 with the highest ΔK value for K = 3, suggesting that three genetic clusters were
321 sufficient to interpret our data (Fig 1).
322

Fig 1. STRUCTURE Harvester results. The most probable number of genetic clusters (K) within the full data set of 410 individuals based on the method described by Evanno et al. (2005) [51]. Delta K graph determined the maximum value at K = 3.

Based on a membership probability threshold of 0.70, 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 (Fig 2A).

Fig 2. Genetic structure of the *Vitis* germplasm accessions obtained on the basis of 17 microsatellite markers. 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. [A] First round of STRUCTURE analysis, inferred genetic structure for K = 3. Cluster 1 (C1): genetic predominance of the species *V. vinifera*; cluster 2 (C2): genetic predominance of the species *V. labrusca*; cluster 3 (C3): genetic predominance of wild *Vitis* species; Admixture: interspecific hybrids with a membership of q < 0.70. [B] Second round of STRUCTURE analysis. WG: wine grape accessions related to *V. vinifera*; TG: table grape accessions related to *V. vinifera*; NG: 'Niagara' accessions; IS: 'Ives' and 'Isabella' accessions; L1 and L2: Others *V. labrusca* hybrids; WV: accessions

347 related to wild *Vitis* species; VR: *V. rotundifolia* accessions; SS: accessions related to
348 the Seibel series; OH: complex interspecific hybrids.
349
350 Of the 304 observed alleles, 227 were shared among the groups; the remaining
351 77 represented private alleles (Pa) in different groups of accessions (Fig 3). The
352 VVMD27 locus had the largest number of private alleles of the 17 SSR markers used in
353 this study (9). Clusters 1 and 2, constituted by accessions related to the most cultivated
354 species of grapevine, *V. vinifera* and *V. labrusca*, respectively, had the smallest number
355 of private alleles (5 and 1, respectively). The largest number of private alleles was found
356 in cluster 3 (53), constituting 72.60% of the total private alleles.
357
358 **Fig 3. Private allele (Pa) frequencies obtained from the genotyping of 410**
359 **grapevine accessions on the basis of 17 microsatellite loci.** X-axis: Private alleles
360 frequencies; Y-axis: groups identified by STRUCTURE analyses at K = 3. The dashed
361 line indicates the cutoff for the occurrence of rare alleles (frequency = 0.05).
362
363 A subsequent round (second round) of STRUCTURE allowed the identification
364 of secondary clusters within the three main genetic clusters (Fig 2B). In Cluster 1, the
365 accessions were divided into two subgroups (K = 2), one formed mainly by wine grapes
366 (WG) (n = 115) and the other by table grapes (VT) (n = 92). This finer-scale clustering
367 divided Cluster 2 into 4 subgroups (K = 4). The NG subgroup (n = 15) was composed of
368 'Niagara' and its mutations. In the IS subgroup (n = 11), the cultivars Ives, Isabella, and
369 Isabella mutations were found. The remaining *V. labrusca* hybrids were allocated to
370 subgroups L1 (n = 18) and L2 (n = 10). In cluster 3, the second round also divided the
371 accessions into two subgroups (K = 2), the *V. rotundifolia* accessions were assigned to

372 the VR subgroup (n = 11), and the others accessions related to wild *Vitis* species were
373 allocated to the WV subgroup (n = 40).

374 Although the Admixture group contained a large number of heterogeneous
375 accessions, a subsequent round of STRUCTURE was also performed on this set to
376 identify possible clustering patterns. As a result, the analysis revealed the presence of
377 two subgroups (K = 2). Accessions of the Seibel series and hybrids including cultivars
378 of this complex in their genealogy were separated from the other hybrids and assigned
379 to the SS subgroup (n = 31). The remaining 67 accessions of the Admixture group were
380 in the OH subgroup.

381 Additionally, DAPC was performed with no prior information about the
382 groupings of the evaluated accessions. Inspection of the BIC values (S1 Fig.) revealed
383 that the division of the accessions into nine clusters was the most likely scheme to
384 explain the variance in this set of accessions. In the preliminary step of data
385 transformation, the maintenance of 120 principal components (PCs) allowed the DAPC
386 to explain 94% of the total genetic variation.

387 Initially, the DAPC scatterplot based on the first and second discriminant
388 functions showed the formation of three main distinct groups, with great genetic
389 differentiation of clusters 8 (dark green) and 9 (green) from the others (Fig 4A). In a
390 subsequent DAPC, outlier clusters 8 and 9 were removed to improve the visualization
391 of the relationship of the other clusters (Fig 4B). In this second scatterplot, clusters 1
392 (magenta) and 7 (purple) showed greater genetic differentiation, with low variance
393 within the groups, as well as no case of overlap with another cluster, indicating a strong
394 genetic structure. The maintenance of 250 principal components (PCs) allowed the
395 second DAPC to explain 100% of the total genetic variation.

396

Fig 4. 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.

The allocation of individuals into clusters according to the DAPC showed several similarities to those achieved in the second round of STRUCTURE, and both analyses showed the same pattern of clustering. Essentially, clusters 1 (magenta), 2 (yellow), 8 (dark green), and 9 (green) of the DAPC reflected the subgroups SS, TG, VR, and WV detected by the STRUCTURE second round, respectively, and the WG subgroup corresponded to DAPC clusters 5 (red) and 3 (orange).

In the case of the *V. labrusca* hybrids, the analyses resulted in a slightly different division. DAPC separated these accessions in clusters 6 (blue) and 7 (purple), basically assigning 'Niagara' accessions in cluster 6 and the other *V. labrusca* hybrids in cluster 7. The STRUCTURE second round also identified 'Niagara' accessions as a separate group (NG); however, a more refined division was performed in the other hybrids, separating them into 3 subgroups. DAPC cluster 4 (brown) did not correspond to any subgroup identified by the STRUCTURE second round; this cluster was formed mostly by hybrids developed by the IAC breeding program used as table grapes.

Finally, we constructed a dendrogram using the neighbor-joining method from the distance matrix based on Rogers' distance to confirm the relationships among the

422 accessions (Fig 5). The dendrogram showed a pattern that was consistent with those
423 from the above-described two analyses. The group formed by the *V. rotundifolia*
424 accessions and the other wild species was clearly separated from the cultivated *Vitis*
425 species, as seen in the DAPC. There was also a strong separation between accessions
426 related to *V. labrusca* and other accessions. The wine grape accessions of *V. vinifera*
427 were mainly concentrated at the top of the dendrogram, while the table grape accessions
428 of this species were found at the bottom. However, the other hybrids (IAC, Seibel
429 series, and others) were scattered among all the groups formed by the dendrogram.
430
431 **Fig 5. Neighbor-joining dendrogram based on Rogers' distance calculated from the**
432 **dataset of 17 microsatellite markers across 410 grapevine accessions.** Accessions
433 colored according to species group.
434

435 ## Validation analysis of molecular profiles

436 The identification of 145 accessions was validated through matches with data
437 available in the literature and databases. The results also confirmed matches to reference
438 profiles of clones based on somatic mutations. Another 42 accessions showed molecular
439 profiles that matched a validated reference profile of a different prime name, indicating
440 mislabeling (S1 Table).

441 The molecular profiles of the remaining 223 accessions did not match any
442 available reference profile. This accession group included wild species and cultivars
443 from grapevine breeding programs in Brazil (the IAC and Embrapa), the United States,
444 and France (Seibel series). The molecular profiles of more than 100 hybrids developed
445 by the IAC were reported for the first time.

446 The accessions '101-14', 'Bailey', 'Black July', 'Carlos', 'Carman', 'Castelão',
447 'Catawba Rosa', 'Elvira', 'Moscatel de Alexandria', and 'Regent' showed a different
448 profile than the reference profile of the same name and did not match any other
449 available reference profile. However, additional morphological and source information
450 is needed to validate their identification. To avoid possible confusion, these accessions
451 were indicated as "Unknown".
452 After correcting the mislabeling, 22 cases of duplicates were identified, all with
453 accessions of the same name and the same molecular profile. Accessions identified with
454 different names but having the same molecular profile were classified as synonyms.
455 Thirty-one synonymous groups were elucidated in this study (S3 Table). Some
456 accessions classified as "Unknown" showed genetic profiles identical to accessions that
457 did not match any available reference profile; examples can be seen in synonymous
458 groups 1, 2, 5, and 6 in S3 Table.
459

460 ## Core collection

461 Three independent sampling proportions were constructed with a size ranging
462 from 10 to 30% of the entire dataset to identify the smallest set of accessions that would
463 be able to represent as much of the available genetic diversity as possible (Table 2).
464 Core 3, composed of 120 accessions, managed to capture 100% of the 304 detected
465 alleles, while the smallest sample (Core 1) managed to capture 243 alleles,
466 approximately 20% less than the total number of alleles detected. The genetic diversity
467 index values obtained for the samples were similar to or higher than those for the entire
468 germplasm. The H_O values ranged from 0.64 (Core 1) to 0.70 (Core 3); the value for
469 Core 3 was similar to that detected for all 410 accessions (0.75). The three samples
470 showed H_E and Ne values higher than those observed for the entire dataset. The H_E

values for all samples 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).

Sample Name	**Size**	**N**	**Na**	**Ne**	**H_O***	**H_E***	**Total SSR diversity captured (%)**
Core 1	0.1	41	243	136.22	0.64 (0.03)	0.85 (0.01)	79.93
Core 2	0.2	82	275	134.53	0.69 (0.04)	0.85 (0.02)	90.46
Core 3	0.3	120	304	137.69	0.70 (0.03)	0.85 (0.01)	100
IAC collection	1.0	410	304	119.42	0.75 (0.03)	0.83 (0.02)	100

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

*Standard error in parentheses.

Core 3 sample was the only one that managed to capture 100% of the alleles, being the best option for use in breeding as a core collection. 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.

We detected a H_E of 0.83 across the entire accession set in the 17 evaluated loci (Table 1). This result is similar to those found in other Brazilian germplasm banks characterized by containing European and American cultivars and an abundance of interspecific hybrids [64,65]. However, this value was higher than that in the Iranian [11] (0.72), Turkish (0.75) [66], and Spanish (0.71) [67] collections, which possessed only *V. vinifera* accessions.

The large number of alleles per locus identified (~18) was likely due to the taxonomic amplitude of the germplasm since a relatively large 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 finding 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 [68].

The high number of alleles obtained by the 17 SSR primer set positively impacted the PIC and discrimination power (D*j*). PIC is an indicator of a marker's informative ability in genetic studies (segregation, population identification, and paternity control), and its value reflects the polymorphism of the marker in the population studied. According to the classification of Botstein et al. (1980) [43], all the loci used can be considered highly informative (PIC > 0.50). The high D*j* values demonstrate that the microsatellite markers used in this study can be considered very effective for grape cultivar discrimination and could be valid to distinguish other accessions that could be introduced into the collection.

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) [69] and Emanuelli et al. (2013) [70]. 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*, *V. labrusca,* and wild *Vitis*. This first structural level is also evidenced in the DAPC analysis and neighbor-joining dendrogram, where it is possible to observe a clear distinction of the accessions associated with *V. labrusca* and wild species.

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 these accessions are crossbreeds between native vine species found in North America such as *V. riparia* Michaux, *V. rupestris* Scheele Michx, and *V. labrusca* L., and a number *V. vinifera* L. cultivars from Europe. The intra- and interspecific crossings carried out during breeding cycles in search of novelties and hybrid vigor promote the

541 miscegenation of grapevine cultivars, resulting in hybrids with a heterogeneous genetic
542 composition.

543 The assignment of these hybrids to groups based on species is often difficult, as
544 these individuals certainly carry alleles from different gene pools, being in an
545 intermediate position and belonging simultaneously to more than one cluster. The
546 accessions 'Campos da Paz' and 'IAC 0457-11 Iracema' are examples of this condition.
547 'Campos da Paz' is an interspecific hybrid resulting from the cross between the
548 cultivated species *V. vinifera* and the wild species *V. ruprestris*. The mixture of two
549 genomes was detected by STRUCTURE, which assigned a membership probability
550 threshold of 0.55 and 0.45 to clusters 1 and 3 respectively, representing the genetic
551 clusters of the two parental species. A similar situation was observed for the accession
552 'IAC 0457-11 Iracema' developed from the cross between the species *V. vinifera* and *V.*
553 *labrusca*, represented by genetic groups 1 and 2, respectively. The hybrid presented an
554 intermediate membership of 0.5 to the two groups. The other accessions from
555 Admixture group exhibited a similar or even more complex origin than these examples,
556 and some of them were derived from crosses between more than three species, having
557 associations with the three clusters simultaneously.

558 Our results demonstrated the largest number of private alleles in cluster 3
559 composed of the wild germplasm (Fig 3). This finding confirms that wild accessions are
560 important reservoirs of genetic variation, with the potential for incorporating new
561 materials into breeding programs in response to the demand for the development of
562 cultivars with different characteristics. Wild grape germplasm is a potential source of
563 unique alleles and provides the breeder with a set of genetic resources that may be
564 useful in the development of cultivars that are resistant to pests and diseases, tolerant to

565 abiotic stresses, and even show enhanced productivity, which makes their conservation
566 of paramount importance [71].

567 The second round of STRUCTURE (Fig 2B) identified similar DAPC clustering
568 patterns (Fig 4), in which the genotypes from *V. vinifera* were separated according to
569 their use. The WG subgroup was composed mainly of wine grapes, such as the
570 accessions 'Syrah', 'Merlot Noir', 'Chenin Blanc', 'Petit Verdot', and 'Cabernet
571 Sauvignon', which showed associations with a membership greater than 0.95,
572 corresponding to DAPC clusters 5 (red) and 3 (orange). The *V. vinifera* accessions of
573 table grapes as 'Centennial Seedless', 'Aigezard', 'Moscatel de Hamburgo', and 'Italia'
574 and their mutations 'Benitaka', 'Rubi', and 'Brazil' were found in the TG subgroup.
575 This subgroup corresponded to cluster 2 (yellow) in the DAPC. In the neighbor-joining
576 dendrogram, the *V. vinifera* accessions were also completely separated in terms of use;
577 the wine grapes were located at the top, and the table grapes were located at the bottom.
578 This result showed that the strong artificial selection based on human usage with wine
579 or table influenced the genetic structure within the cultivated compartment of grapevine,
580 as previously identified in previous studies [69,72].

581 In the DAPC and neighbor-joining dendrogram, two groups were differentiated
582 to a greater extent than the others (Fig 4 and Fig 5), with these groups being formed
583 mainly of wild grapes that are often used as rootstocks. This phenomenon likely
584 occurred because few rootstocks used worldwide contain part of the *V. vinifera* genome
585 [9], while practically all table and wine grape hybrids present in this germplasm contain
586 a part of it. In DAPC analyses, the *V. rotundifolia* accessions constituted the most
587 divergent group. This species is the only one in the germplasm belonging to the
588 *Muscadinia* subgenus, which contains plants with 2n = 40 chromosomes, while the
589 others belong to the *Euvitis* subgenus, with 2n = 38 chromosomes. A high genetic

590 divergence between *V. rotundifolia* and the species in the *Euvitis* subgenus was also
591 observed by Costa et al. (2017) [73] through the use of RAPD molecular markers and
592 by Miller et al. (2013) [74] through SNPs. The species *V. rotundifolia* is resistant to
593 several grapevine pests and diseases [75] and is an important source of genetic material
594 in the development of cultivars and rootstocks adapted to the most diverse
595 environmental conditions and with tolerance and/or resistance to biotic and abiotic
596 factors.

597 The DAPC cluster 1 was formed by only accessions of the Seibel series and
598 hybrids with varieties of this series in their genealogy. The Seibel series is in fact a
599 generic term that refers to several hybrid grapes developed in France at the end of the
600 19th century by Albert Seibel from crosses between European *V. vinifera* varieties and
601 wild American *Vitis* species to develop phylloxera-resistant cultivars with
602 characteristics of fine European grapes [76]. As these hybrids are derived from crosses
603 among three or more species, most of them were identified as Admixture in the first
604 round of the STRUCTURE. A second round of STRUCTURE was carried out in the
605 Admixture group to confirm the structure of these accessions as shown by DAPC. As a
606 result, the Seibel series accessions were separated from the other hybrids to form a
607 subgroup, confirming the existence of a distinct gene pool. The combinations of alleles
608 of different *Vitis* species clearly created unique genetic pools, with many related
609 accessions, since they were developed using the same breeding program, which explains
610 the grouping and genetic distinction.

611 The *V. labrusca* hybrids formed distinct groups in the three analyses. In the
612 DAPC, this accession group was subdivided into two clusters (6 and 7) indicating the
613 presence of a secondary structure between them (Fig 4). Cluster 6 contained only table
614 grape cultivars, including ‘Eumelan’, ‘Niabell’, ‘Highland’, ‘Niagara’, and their

615 mutations, while grape cultivars for processing, including 'Isabella', 'Ives', and
616 'Concord Precoce' were included in cluster 7. In the STRUCTURE second round, these
617 accessions had a more pronounced division (Fig 2B), and the cultivars Niagara and
618 Isabella together with their mutations were assigned to subgroups NG and IS,
619 respectively, while the other accessions were distributed between subgroups L1 and L2.
620 This refined secondary structuring was probably due to the hierarchical STRUCTURE
621 method, since the sensitivity of the program is increased when using a primary cluster in
622 isolation that allows for more detailed subdivisions [49].

623 The IAC breeding program started in 1943 with the aim of obtaining varieties of
624 wine grapes, table grapes, and rootstocks. The first introductions in the Germplasm
625 Bank constituted *V. viniferas* cultivars and Seilbel series hybrids originating in France.
626 Subsequently, wild species and *V. labrusca* hybrids from North America were
627 introduced. Varieties developed around the world continued to be introduced into the
628 IAC germplasm (Table S1) over time, which currently has a large number of accesses
629 originating mainly from the United States, France, and Italy, which correspond to
630 19.51%, 18.04%, and 8.78% of the germplasm, respectively. In smaller quantities,
631 varieties from Argentina, Germany, Armenia, Spain, Japan, Portugal, and other
632 countries are also found.

633 Many of the *V. vinifera* cultivars of the IAC germplasm originating in France,
634 Italy, and Spain are common among grapevine germplasms worldwide, and their use in
635 other studies of genetic diversity has been reported [67,69,77–80]. The American and
636 Brazilian hybrids present in the germplasm are more restricted to collections in North
637 and South America, being rarely reported in European studies [10,65,68].

638 With the results of the first crosses in the IAC breeding program, the hybrids
639 with outstanding characteristics started to be used as parents [81]. Since the beginning

of the program, more than 2,000 crosses have been performed over 50 years, using more than 800 parents [82]. Currently the germplasm has 134 accessions developed from these crossings, corresponding to 32.70% of the entire germplasm. Most of these hybrids are exclusive to this germplasm, and the molecular profiles of 109 are described for the first time in this study.

The broad genetic base and different objectives of the IAC breeding program were responsible for the development of hybrids with a wide genetic diversity, as evidenced in the three analyses revealing IAC hybrids in practically all the clusters. In the dendrogram, the IAC hybrids were highlighted to facilitate this perception (Fig 5). Over time, there has been a decrease in the importance of the wine industry in the State of São Paulo, and the search for table grape varieties has become predominant [82]. Some of these table grape hybrids developed by the IAC formed cluster 4 of the DAPC. The clustering of these hybrids is similar to the case of the Seibel series accessions, where the combinations of alleles from different crossings were probably responsible for the creation of a unique gene pool.

The analyses grouped most of the hybrids with one of their parents; however, cases in which the hybrids were not grouped with any of the parents occurred. Hybrids originating from the same crosses were not always grouped with the same parent. For example, the hybrids ‘IAC 0871-41 Patrícia’ and ‘IAC 0871-13 A Dona’ both resulted from the same crossing between hybrids ‘IAC 0501-06 Soraya’ and ‘IAC 0544-14’ located in DAPC clusters 2 and 3, respectively, hybrid ‘IAC 0871-41 Patrícia’ was grouped with its parent ‘IAC 0501-06 Soraya’, while ‘IAC 0871-13 A Dona’ was grouped with ‘IAC 0544-14’. These findings are easily explained when we consider the genetic biology of the grapevine. In general, grapevine cultivars are highly heterozygous, and crossing between divergent parents results in a highly segregating

665 progeny. In the same progeny, the hybrids are heterogeneous, and they can present
666 characteristics similar to both parents, similar to only one parent, or even different from
667 both parents [83].

668 Since many of the accessions were introduced from different parts of the world
669 and some others have a complex pedigree, it can be difficult to determine their true
670 relationship. In the absence of information on the genetic relationships among most
671 genotypes, it is not possible to determine the most accurate method of grouping.
672 Although the use of multivariate techniques in the recognition of genetic diversity
673 imposes a certain degree of structure in the data, and it is important to use different
674 grouping criteria and the correct structure resulting from most of them to ensure that the
675 obtained result is not an artifact of the technique used. The use of more than one
676 clustering method, due to differences in hierarchization, optimization, and ordering of
677 groups allows the classification to be complemented according to the criteria utilized by
678 each technique and prevents erroneous inferences from being adopted in the allocation
679 of materials within a given subgroup of genotypes [84].

680 The STRUCTURE grouping method could be contested because human
681 manipulation of cultivars (displacements, breeding, clonal propagation) can generate a
682 deviation from Hardy-Weinberg equilibrium; however, in our study, STRUCTURE
683 analysis provided a very consistent attribution of genotypes to clusters. The Admixture
684 group reflects the crossing among genotypes of the three groups identified in the first
685 round of STRUCTURE corresponding to breeding activities in search of novelties and
686 hybrid vigor. Furthermore, this analysis provides important information regarding the
687 genetic composition of the hybrids, providing information about the proportion of each
688 species in their genome. The three primary genetic groups of STRUCTURE were easily
689 distinguished in the other analyses; however, in the DAPC analysis, new levels of

690 structure were revealed within these primary groups. The DAPC analysis also provided
691 information about the genetic divergence between the clusters, allowing the
692 identification of related ones.

693 The STRUCTURE second round was carried out to investigate the presence of
694 subclusters within the primary clusters and simultaneously validate the levels of
695 structure obtained in the DAPC analysis. Most of the subgroups found in the
696 STRUCTURE second round corresponded to the division obtained by the DAPC
697 analysis, although some structural levels were different. These differences between
698 analyses do not invalidate their results but rather bring complementary information that
699 enhances understanding of the genetic structure and genetic relationship of germplasm
700 accessions. The grouping based on the species of accessions was also evidenced by
701 neighbor-joining dendrogram, but the differential of this analysis further provided visual
702 information on the genetic relationship of the accessions within the groups. In the
703 dendrogram, the genetic distance between two specific individuals was easily verified,
704 providing a useful tool in breeding programs, mainly for the selection of divergent
705 parents.

706 The information obtained by the STRUCTURE, DAPC, and neighbor-joining
707 dendrogram provides important knowledge for the management of germplasm diversity.
708 The identification of divergent groups guides crossings in breeding programs,
709 facilitating the appropriate combination of parents to obtain progeny with wide genetic
710 variability, allowing the maximization of heterosis and making it possible to obtain
711 individuals with superior characteristics. Information about the available genetic
712 diversity is valuable because if properly explored, it can reduce vulnerability to genetic
713 erosion through the avoidance of crosses between genetically related genotypes while

714 also accelerating genetic progress related to characteristics of importance to grape
715 growth [85].
716

717 Identification analysis: misnamed and synonymous cases

718 Considering the vast diversity of names for the different varieties of grapevine,
719 standardization is necessary. Errors due to homonyms, synonyms, differences in
720 spelling, and misnamed accessions impede estimation of the real number of different
721 accessions that are present in grapevine collections, with a negative impact on grapevine
722 breeding programs. Therefore, the verification of true-to-type accessions is
723 indispensable [33]. For grapevine, a 7-SSR genotyping system has been established as a
724 useful tool for identification and parentage analysis, allowing the allele length of
725 varieties to be comparably scored by different institutions [34,62].

726 In this study, 42 cases of misnaming were found by comparing the molecular
727 profiles obtained with the information available in the literature and databases (S1
728 Table). The molecular profile of the accession 'Cabernet Franc' corresponded to the
729 cultivar Merlot Noir, and the molecular profile of the accession 'Merlot Noir'
730 corresponded to the cultivar Cabernet Franc, clearly indicating an exchange of
731 nomenclature between these accessions. 'Cabernet Franc' is one of the parents of
732 'Merlot Noir', and some morphological traits of these two cultivars are quite similar
733 [86], which certainly contributed to the occurrence of this mistake.

734 The molecular profile of the accession 'Magoon' matched that of 'Regale' in the
735 present study, and 'Regale' had a similar molecular profile to those obtained by Schuck
736 et al. (2011) [87] and Riaz et al. (2008) [88], indicating that 'Magoon' was misnamed at
737 the time of introduction and that both accessions were the cultivar Regale. In Brazil, the
738 same case of misnaming was also reported by Schuck et al. (2011) [87]. A misspelling

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

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

750 Additionally, 31 synonymous groups were identified (S3 Table). Cases of
751 synonymy could correspond to clones of the same cultivar that show phenotypic
752 differences due to the occurrence of somatic mutations [89,90]. Mobile elements are
753 known to generate somatic variation in vegetatively propagated plants such as
754 grapevines [91,92]. Carrier et al. (2012) [91] observed that insertion polymorphism
755 caused by mobile elements is the major cause of mutational events related to clonal
756 variation. In grape, retrotransposon-induced insertion into *VvmybA1*, a homolog of
757 *VlmybA1-1*, is the molecular basis of the loss of pigmentation in a white grape cultivar
758 of *V. vinifera* due to the lack of anthocyanin production [93].

759 The detection of somatic mutations is very difficult with a small number of
760 microsatellite markers, especially when they are located in noncoding regions of the
761 genome [94]. This was the case for synonymous groups 19, 20, 21, 22, in S3 Table,
762 such as the cultivar Italia and its mutations 'Rubi', 'Benitaka', and 'Brasil', which differ
763 in terms of the color of berries, with white, pink, red, and black fruits, respectively, and

764 are cultivated as distinct cultivars in Brazil. This was also the case for 'Pinot Gris', a
765 variant with gray berries arising from 'Pinot Noir', which has black berries. The
766 mutations that occurred in the cultivar Niagara can be distinguished in terms of the
767 color, size, and shape of the berries, and they may even lead to a lack of seeds, such as
768 the apyrenic accession 'Niagara Seedless' or 'Rosinha' [95].

769 The accession 'Tinta Roriz' was identified as a synonym of the cultivar
770 Tempranillo Tinto in this study; this synonym is already registered in the VIVC and is
771 widely used in regions of Portugal [96]. The wild species *V. doaniana* and *V.*
772 *berlandieri* have the same molecular profile, indicating a case of mislabeling; certainly,
773 some mistake was made during the acquisition of these materials, and the same
774 genotype was propagated with different names.

775 The occurrence of misidentification is common, especially for old clonal species
776 such as *Vitis* spp., and it can occur during any stage of accession introduction and
777 maintenance. It has been observed that 5 to 10% of the grape cultivars maintained in
778 grape collections are incorrectly identified [97,98]. In a new place, a certain genotype
779 may receive a new name, confusing samples and the maintenance of accessions in
780 germplasm banks [99]. The correct identification of accessions is fundamental to
781 optimize germplasm management and for the use of germplasm in ongoing breeding
782 programs since related genotypes will not be chosen for field experiments or controlled
783 crosses. The identification of the existence of synonyms, homonyms, and misnamed
784 accessions is essential to prevent future propagation and breeding errors [87] and in
785 helping to reduce germplasm maintenance costs without the risk of losing valuable
786 genetic resources. Since morphological descriptors are highly influenced by
787 environmental factors, molecular analyses can support identification. SSR markers have
788 often been considered very efficient at the cultivar level since they can be easily used to

789 distinguish different cultivars; however, they are less effective in differentiating clones
790 [9]. The results of molecular analysis should not replace ampelographic observations
791 but should be integrated with such observations, mainly for the identification of somatic
792 mutations.

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

801

802 ## Development of a core collection

803 The intention for the development of a core collection is to represent the genetic
804 diversity of the entire germplasm in a reduced set of accessions that is feasible to
805 handle. The efficiency of the approach based on SSR profiles in identifying a core
806 collection was already demonstrated for grapevine by Le Cunff et al. (2008) [100],
807 Cipriani et al. (2010) [101], Emanuelli et al. (2013) [70] and Migliaro et al. (2019) [29].

808 In this study, 120 accessions (Core 3) were necessary to capture all the allelic
809 diversity of the whole collection, which is equivalent to approximately 30% of all
810 accessions (Table 2). In *V. vinifera* subsp. *sativa* core collections, the same result was
811 obtained with smaller percentages of individuals, from 4 to 15% [70,100,101].
812 According to Le Cunff et al. (2008) [100], the use of only cultivated genotypes of *V.*
813 *vinifera* subsp. *sativa* is one of the reasons for the small number of individuals in the

core collection since cultivated genotypes tend to be less diverse than wild counterparts [12,102].

Migliaro et al. (2019) [29] 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 [100].

Among the 120 genotypes in Core 3, 82 were identified as interspecific hybrids, with 13 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

838 useful for the development of new breeding strategies, future phenotyping efforts, and
839 genome-wide association studies.
840

841 # Conclusions

842 A wide range of genetic diversity was revealed in the studied germplasm,
843 ensuring the conservation of a large portion of grapevine genetic resources. The genetic
844 diversity showed a pattern of structuring based on the species and use of accessions, as
845 evidenced in a manner similar to the three structuring analyses. In addition, each of the
846 analyses provided different information that was be complementary and equally
847 valuable for breeding.

848 Taken together, our results can be used to efficiently guide future breeding
849 efforts, whether through traditional hybridization or new breeding technologies. The
850 obtained information may also enhance the management of grapevine germplasms and
851 provide molecular data from a large set of genetic resources that contribute to
852 expanding existing database information.
853

854 # Acknowledgments

855 The authors are grateful to the São Paulo Research Foundation (FAPESP) and
856 Coordination for the Improvement of Higher Education Personnel (CAPES) for
857 supporting the project and researchers.
858

859 # References

860 1. Torregrosa L, Vialet S, Adivèze A, Iocco-Corena P, Thomas MR. Grapevine
861 (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

887 management. Theor Appl Genet. 2011;122: 1233–1245. doi:10.1007/s00122-
888 010-1527-y
889 10. Leão PCS, Riaz S, Graziani R, Dangl GS, Motoike SY, Walker MA.
890 Characterization of a Brazilian grape germplasm collection using microsatellite
891 markers. Am J Enol Vitic. 2009;60: 517–524.
892 11. Doulati-Baneh H, Mohammadi SA, Labra M. Genetic structure and diversity
893 analysis in Vitis vinifera L. cultivars from Iran using SSR markers. Sci Hortic
894 (Amsterdam). 2013;160: 29–36. doi:10.1016/j.scienta.2013.05.029
895 12. De Souza LM, Guen V Le, Cerqueira-Silva CBM, Silva CC, Mantello CC,
896 Conson ARO, et al. Genetic diversity strategy for the management and use of
897 rubber genetic resources: More than 1,000 wild and cultivated accessions in a
898 100-genotype core collection. PLoS One. 2015;10: 1–20.
899 doi:10.1371/journal.pone.0134607
900 13. Mohammadi SA, Prasanna BM. Analysis of genetic diversity in crop plants -
901 Salient statistical tools and considerations. Crop Sci. 2003;43: 1235–1248.
902 doi:10.2135/cropsci2003.1235
903 14. Boursiquot JM, This P. Les nouvelles techniques utilisées en ampélographie:
904 informatique et marquage. J Int des Sci la Vigne du Vin. 1996;Special: 13–23.
905 15. Roychowdhury R, Taoutaou A, Hakeem KR, Ragab M, Gawwad A, Tah J.
906 Molecular Marker-Assisted Technologies for Crop Improvement. 2014; 241–258.
907 doi:10.13140/RG.2.1.2822.2560
908 16. Jarne P, Lagoda PJL. Microsatellites, from molecules to populations and back.
909 Trends Ecol Evol. 1996;11: 424–429. doi:10.1016/0169-5347(96)10049-5
910 17. Lamboy WF, Alpha CG. Using simple sequence repeats (SSRs) for DNA
911 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. 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

30. Zarouri B, Vargas AM, Gaforio L, Aller M, de Andrés MT, Cabezas JA. Whole-genome genotyping of grape using a panel of microsatellite multiplex PCRs. Tree Genet Genomes. 2015;11. doi:10.1007/s11295-015-0843-4

31. Popescu CF, Maul E, Dejeu LC, Dinu D, Gheorge RN, Laucou V, et al. Identification and characterization of Romanian grapevine genetic resources. Vitis - J Grapevine Res. 2017;56: 173–180. doi:10.5073/vitis.2017.56.173-180

32. 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

33. 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

34. 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.

35. 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

36. 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

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

38. 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

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

40. 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

41. 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

42. 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

43. 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.

44. 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

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

46. Summers K, Amos W. Behavioral, ecological, and molecular genetic analyses of reproductive strategies in the Amazonian dart-poison frog, Dendrobates ventrimaculatus. Behav Ecol. 1997;8: 260–267. doi:10.1093/beheco/8.3.260

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

48. Porras-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo Á, Lareu M V. An overview of STRUCTURE: Applications, parameter settings, and supporting software. Front Genet. 2013;4: 1–13. doi:10.3389/fgene.2013.00098

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

50. Pritchard JK, Wen X, Falush D. Documentation for structure software: Version 2.2. Chicago, USA: University of Chicago; 2007.

51. 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

52. 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

53. 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

54. 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

55. 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.

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

57. 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

58. 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

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

60. 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

61. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC Genet. 2010;11. doi:10.1186/1471-2156-11-94

62. 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

63. 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

64. 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

65. 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

66. Boz Y, Bakir M, Çelikkol BP, Kazan K, Yilmaz F, Çakir B, et al. Genetic

1062 characterization of grape (Vitis vinifera L.) germplasm from Southeast Anatolia
1063 by SSR markers. Vitis - J Grapevine Res. 2011;50: 99–106.
1064 67. Ibáñez J, De Andrés MT, Molino A, Borrego J. Genetic study of key Spanish
1065 grapevine varieties using microsatellite analysis. Am J Enol Vitic. 2003;54: 22–
1066 30.
1067 68. Pollefeys P, Bousquet J. Molecular genetic diversity of the French – American
1068 grapevine hybrids cultivated in North America. Genome. 2003;46: 1037–1048.
1069 doi:10.1139/g03-076
1070 69. Laucou V, Launay A, Bacilieri R, Lacombe T, Andre MT De, Hausmann L, et al.
1071 Extended diversity analysis of cultivated grapevine Vitis vinifera with 10K
1072 genome-wide SNPs. PLoS One. 2018;13: 1–27.
1073 doi:https://doi.org/10.1371/journal.pone.0192540
1074 70. Emanuelli F, Lorenzi S, Grzeskowiak L, Catalano V, Stefanini M, Troggio M, et
1075 al. Genetic diversity and population structure assessed by SSR and SNP markers
1076 in a large germplasm collection of grape. BMC Plant Biol. 2013;13: 1–17.
1077 doi:10.1186/1471-2229-13-39
1078 71. Hajjar R, Hodgkin T. The use of wild relatives in crop improvement : A survey
1079 of developments over the last 20 years The use of wild relatives in crop
1080 improvement : A survey of developments over the last 20 years. Euphytica.
1081 2007;156: 1–13. doi:10.1007/s10681-007-9363-0
1082 72. Aradhya MK, Dangl GS, Prins BH, Boursiquot JM, Walker MA, Meredith CP, et
1083 al. Genetic structure and differentiation in cultivated grape, Vitis vinifera L.
1084 Genet Res. 2003;81: 179–192. doi:10.1017/S0016672303006177
1085 73. Costa AF, Teodoro PE, Bhering LL, Tardin FD, Daher RF. Molecular analysis of
1086 genetic diversity among vine accessions using DNA markers. Genet Mol Res.

2017;16: 1–9. doi:10.4238/gmr16029586

74. 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

75. 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

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

77. Sefc KM, Lopes MS, Lefort F, Botta R, Roubelakis-Angelakis KA, Ibáñez J, et al. Microsatellite variability in grapevine cultivars from different European regions and evaluation of assignment testing to assess the geographic origin of cultivars. Theor Appl Genet. 2000;100: 498–505. doi:10.1007/s001220050065

78. Crespan M, Fabbro A, Giannetto S, Meneghetti S, Petrussi C, Del Zan F, et al. Recognition and genotyping of minor germplasm of Friuli Venezia Giulia revealed high diversity. Vitis - J Grapevine Res. 2011;50: 21–28.

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

80. Schneider A, Raimondi S, Pirolo CS, Marinoni DT, Ruffa P, Venerito P, et al. Genetic characterization of grape cultivars from Apulia (southern Italy) and synonymies in other Mediterranean regions. Am J Enol Vitic. 2014;65: 244–249.

doi:10.5344/ajev.2013.13082

81. Pommer C V. Uva. In: Furlani AMC, editor. O melhoramento de plantas no Instituto Agronômico. Campinas: Instituto Agronômico; 1993. pp. 489–524.

82. Ferri CP, Pommer CV. Quarenta e oito anos de melhoramento da videira em São Paulo, Brasil. Sci Agric. 1995;52: 107–122. doi:10.1590/s0103-90161995000100019

83. Eibach R, Töpfer R. Traditional grapevine breeding techniques. Grapevine Breeding Programs for the Wine Industry. Elsevier Ltd; 2015. doi:10.1016/B978-1-78242-075-0.00001-6

84. Arriel NHC, Di Mauro AO, Di Mauro SMZ, Bakke OA, Unêda-Trevisoli SH, Costa MM, et al. Técnicas multivariadas na determinação da diversidade genética em gergelim usando marcadores RAPD. Pesqui Agropecuária Bras. 2006;41: 801–809. doi:10.1590/s0100-204x2006000500012

85. 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

86. 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

87. 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

88. 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

89. 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

90. 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

91. 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

92. 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

93. Kobayashi S, Goto-Yamamoto N, Hirochika H. Retrotransposon-Induced Mutations in Grape Skin Color. Science (80-). 2004;304: 982. doi:10.1126/science.1095011

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

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

96. 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

97. 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.

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

99. 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

100. 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

101. 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

102. 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

103. Adam-Blondon AF, Roux C, Claux D, Butterlin G, Merdinoglu D, This P. Mapping 245 SSR markers on the Vitis vinifera genome: A tool for grape

genetics. Theor Appl Genet. 2004;109: 1017–1027. doi:10.1007/s00122-004-1704-y

Supporting Information

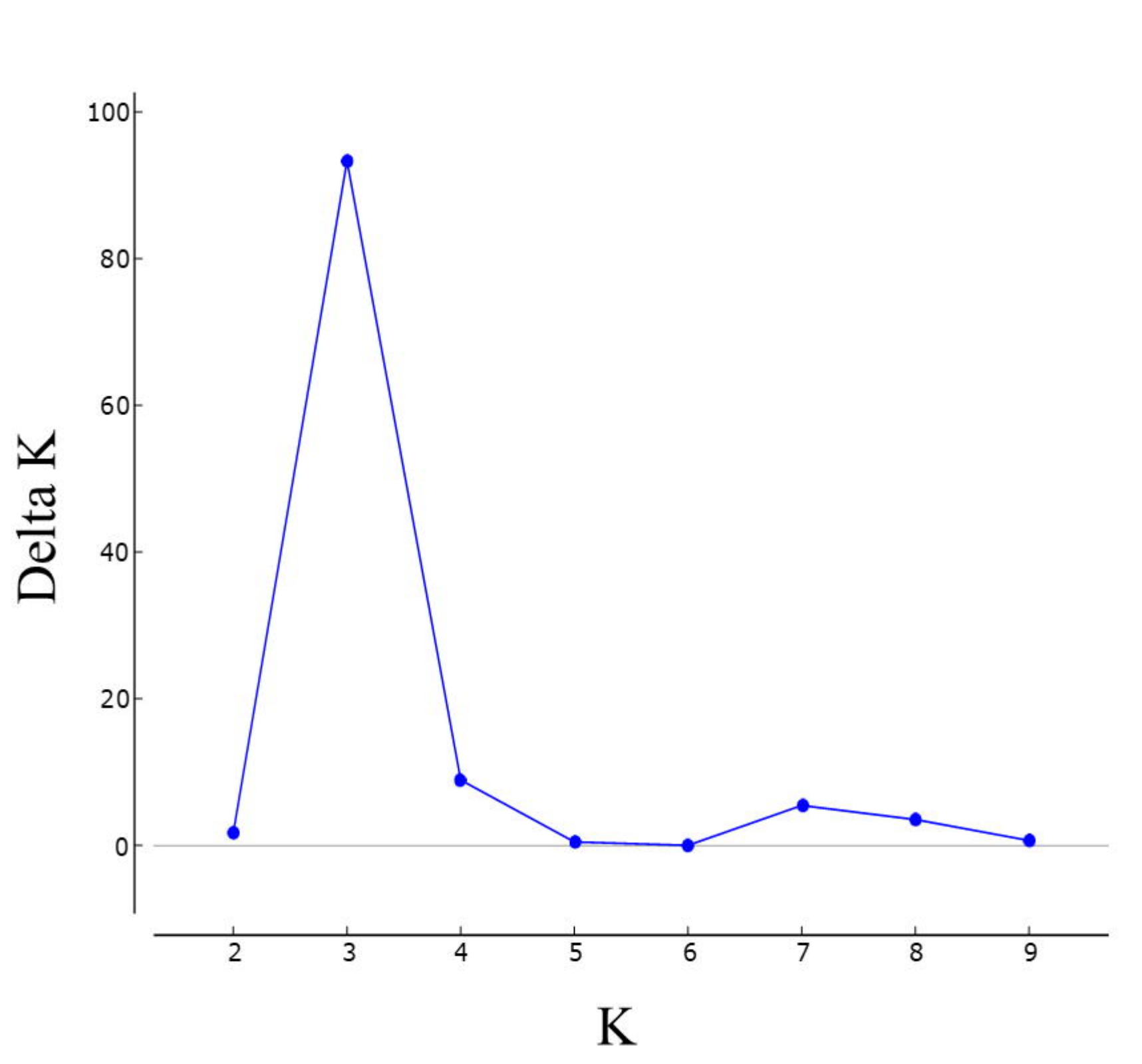
S1 Table. Genotypes used in this study. Detailed characteristics: SSR matches, usage, country of origin, species, group assignments according to the DAPC, and STRUCTURE first and second round, core collection composition, and genotyping results obtained with the 17 microsatellite markers.

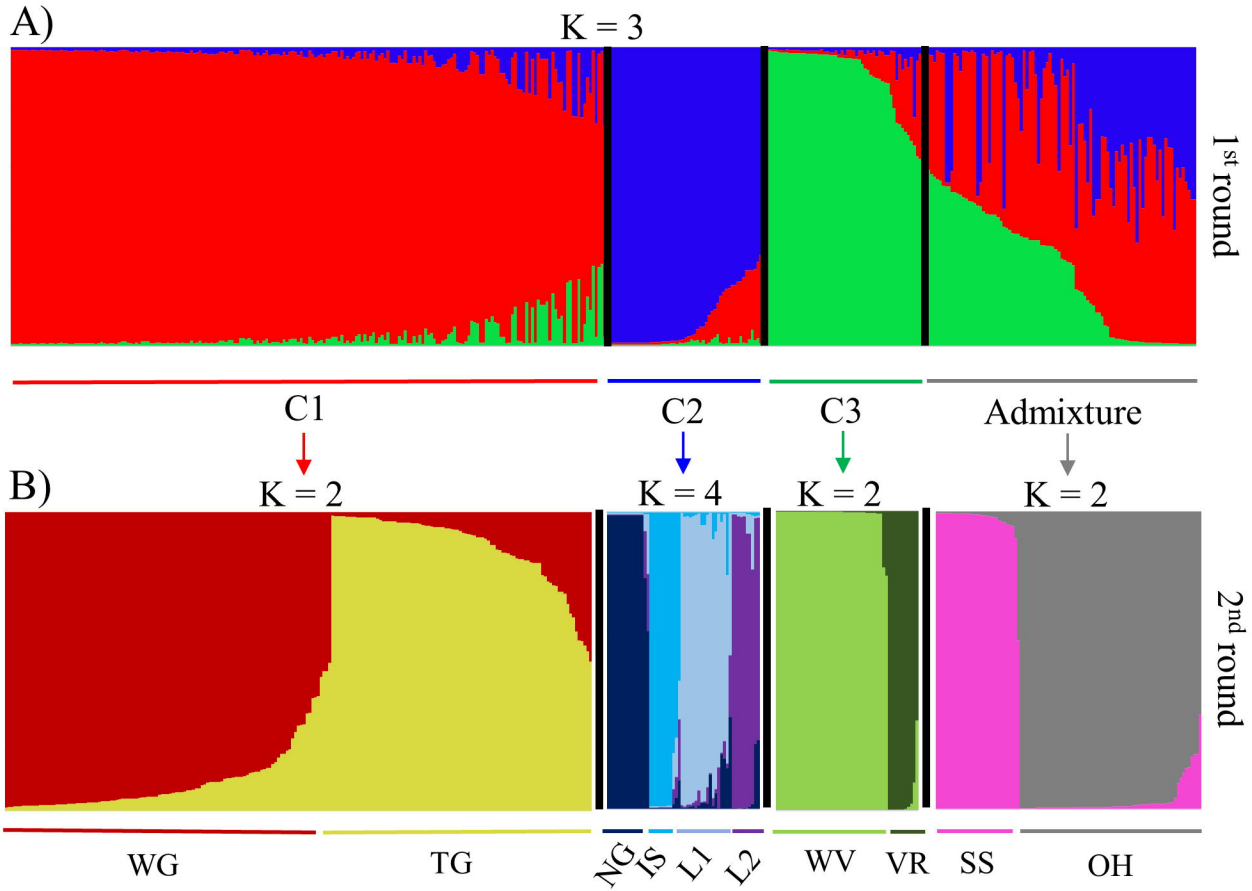
S2 Table. Name, linkage group, microsatellite sequences, and references of the SSR markers used in this study.

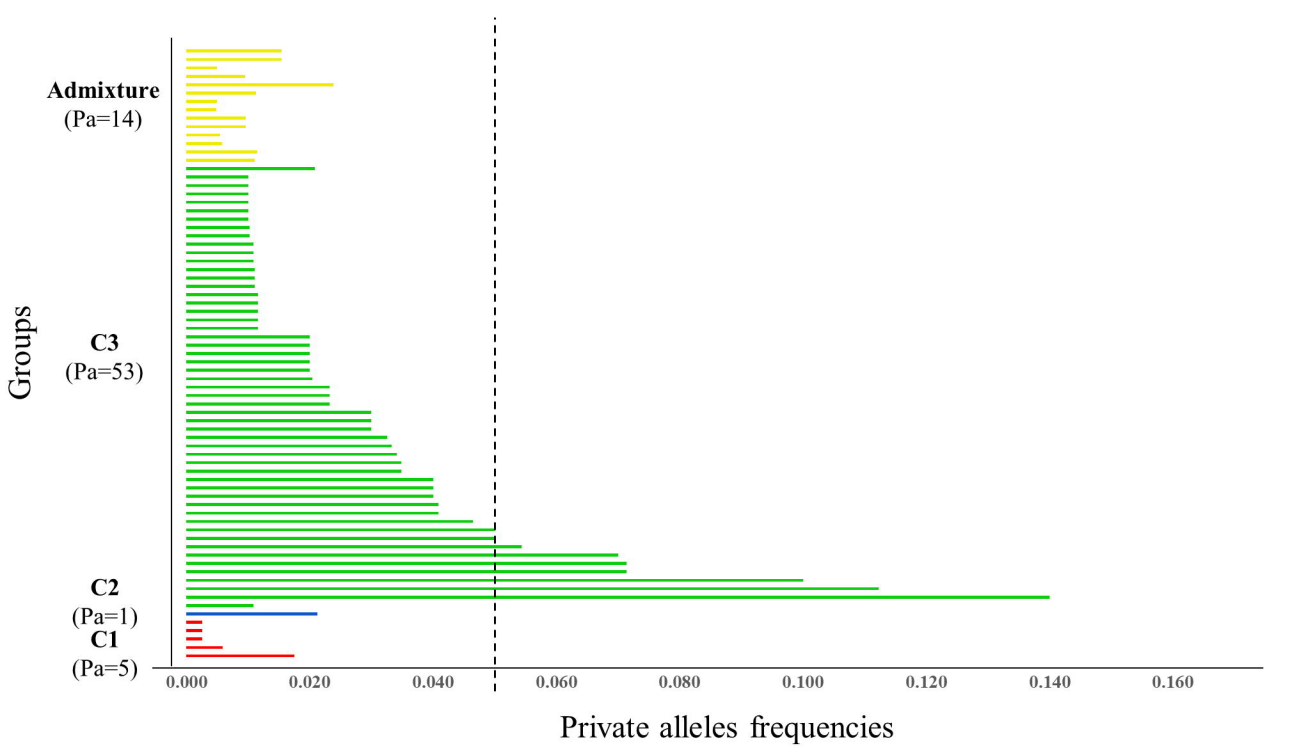
S3 Table. List of synonyms found in the *Vitis* spp. Germplasm Bank of the Agronomic Institute of Campinas (IAC) by SSR analysis.

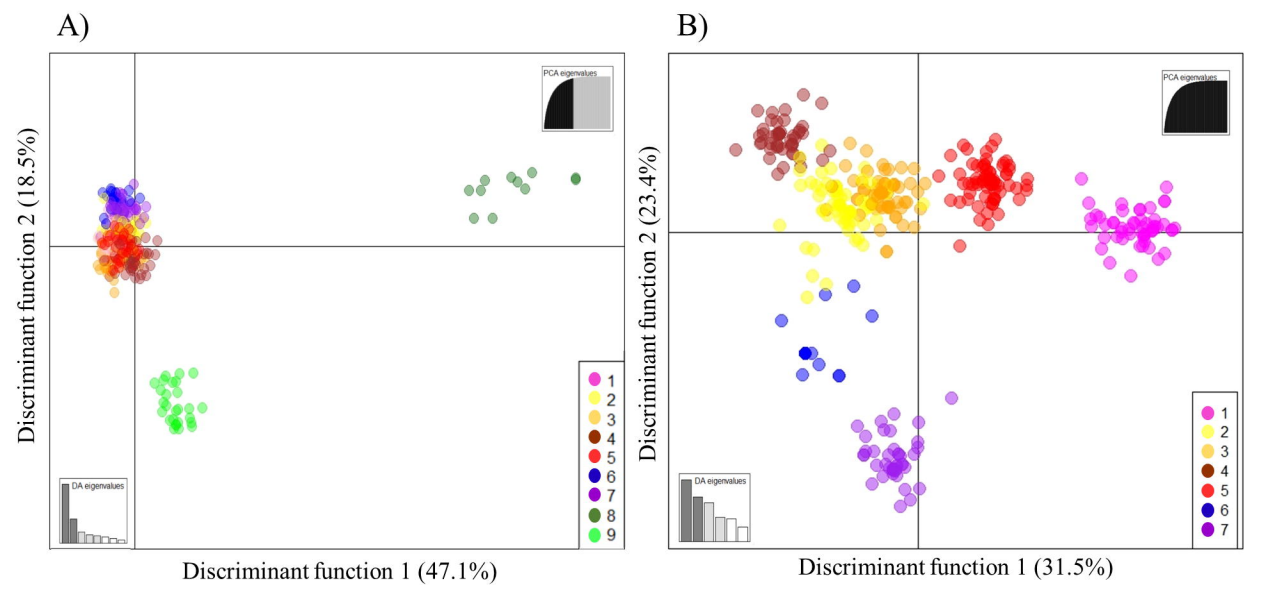
S1 Fig. Bayesian information criterion (BIC) values for different numbers of clusters. The accepted true number of clusters was nine.

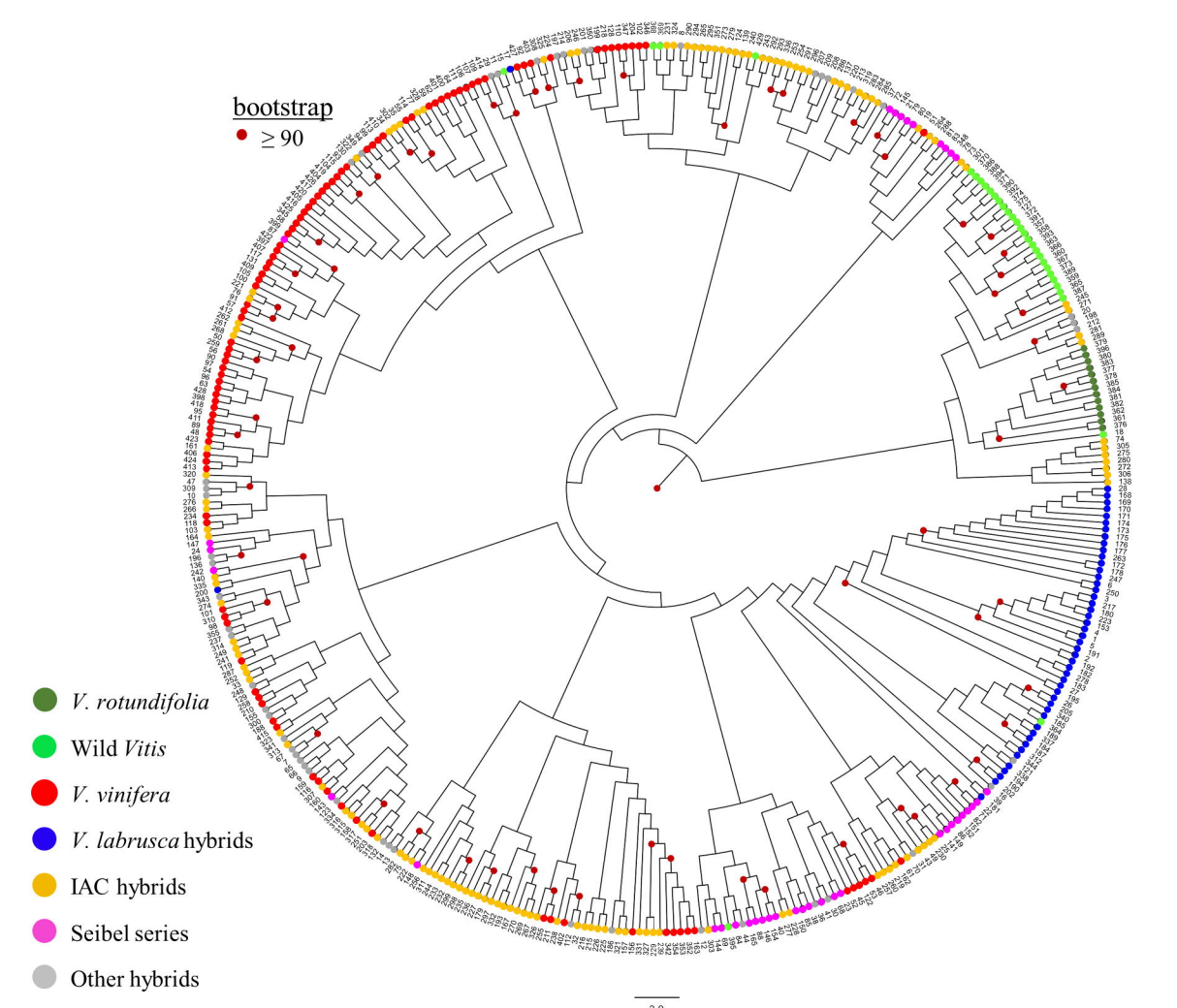
S2 Fig. Harvester results for STRUCTURE second round. Graphics for the detection of the most probable number of groups (K) estimated based on the method described by Evanno et al. (2005) [51]. [A] Cluster 1 - Highest peak for K□=□2. [B] Cluster 2 - Highest peak for K□=□4. [C] Cluster 3 - Highest peak for K□=□2. [D] Admixture group - Highest peak for K□=□2.


Delta K
100
80
60
40
20
0
2
3
4
5
6
7
8
9
K


A)
K = 3
1st round
C1
C2
C3
Admixture
B)
K = 2
K = 4
K = 2
K = 2
2nd round
WG
TG
NG
IS
L1
L2
WV
VR
SS
OH


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


A)
PCA eigenvalues
DA eigenvalues
1
2
3
4
5
6
7
8
9
Discriminant function 2 (18.5%)
Discriminant function 1 (47.1%)
B)
PCA eigenvalues
DA eigenvalues
1
2
3
4
5
6
7
Discriminant function 2 (23.4%)
Discriminant function 1 (31.5%)


bootstrap
≥ 90
V. rotundifolia
Wild Vitis
V. vinifera
V. labrusca hybrids
IAC hybrids
Seibel series
Other hybrids