Grapevine non-*vinifera* genetic diversity assessed by SSR markers as a starting-point for new rootstock breeding programs

Daniele Migliaro[1], Gabriella De Lorenzis[2], Giovambattista Simone Di Lorenzo[2], Barbara De Nardi[1], Massimo Gardiman[1], Osvaldo Failla[2], Lucio Brancadoro[2], Manna Crespan[1*]

[1]CREA - Research Centre for Viticulture and Enology, Viale XXVIII Aprile, 26 – 31015 Conegliano (Treviso), Italy

[2]Department of Agricultural and Environmental Sciences, Via Celoria, 2 – 20133 Milano, Italy

[*]Corresponding author, email: manna.crespan@crea.gov.it

Acknowledgements

We are grateful to Dr. Vito Falco (Vivai Paulsen of Palermo, Italy) for his contribution in the collection of plant materials. This research was funded by Winegraft srl and SIV (Service for grapevine identification of CREA Viticulture and Enology of Conegliano).

Abstract

Microsatellite markers are valuable tool to facilitate management of germplasm collections and to access the genetic diversity. In this study, the genetic characterization of a large collection of 379 rootstocks and other non-*vinifera* accessions maintained at the University of Milan (Italy) has been reported. Accessions were genotyped with 22

highly polymorphic microsatellite markers (including the nine 'international' loci used for grapevine identification, three VMC, three VrZAG and seven VChr loci), but only 17 loci were retained for cultivar identification, to investigate genetic diversity, for pedigree analysis, to infer population structure and to design a core collection. Two hundred thirty-two unique genotypes were identified. The allelic profiles of sixty-nine rootstocks were confirmed according to literature and databases, while the profiles of 44 rootstocks were proposed for the first time. Pedigree analysis highlighted 77 parent-offspring (PO) trios and XX PO relationships, some of them already known and some other new. Analysis of the genetic structure showed a more likely number of three ancestral groups, with a high percentage of admixed samples. A structure based on the genetic background of genotypes was not observed. A core collection of seventy genotypes captured 100% of the entire number (373) of detected alleles. Most of these genotypes were unidentified or poorly characterized genotypes. The information provided in this paper could assist breeders in better addressing their efforts in the exploitation of still unexplored individuals useful for long-term breeding plans.

Keywords: core collection, North American *Vitis* species, VChr microsatellites, population structure, VVMD25 and VVMD32.

Introduction

Until the mid-19th century, European vineyards consisted of own-rooted *Vitis vinifera* vines. The inadvertent introduction of phylloxera (*Daktulosphaira vitifoliae*; Fitch 1855), a phytophage that attacks *V. vinifera* roots, by importing American vines, nearly

destroyed the European wine industry and modern viticulture became dependent on North American *Vitis* species (Snyder 1937, Granett et al. 2001). American grape species evolved in the presence of phylloxera, their leaves and roots are colonized by this insect but both organs tolerate its presence without harmful damage. The agronomical practice of grafting *V. vinifera* cultivars onto American species and their hybrids is used to avoid phylloxera heavy damage. Despite the grape industry relying on a very small number of well-characterized rootstocks (Zavaglia et al. 2016), there are no estimates for the exact number of rootstock cultivars existing today. Ampelography and ampelometry, based on morphological differences, are traditional methods for grapevine cultivar identification. Unfortunately, ampelographic criteria are not satisfactory to distinguish among rootstocks, because of their common parentage from *V. riparia*, *V. rupestris* and *V. berlandieri* (Poczai et al. 2013). Nowadays, identification could be assisted by Simple Sequence Repeats (SSR) or microsatellite markers (Maul et al. 2012). In the last twenty years, microsatellites became the preferential and most reliable tool for grapevine cultivar identification. Much work has been done so far on wine and table varieties, whereas data about genotyping of rootstocks are poor (Lin and Walker 1998, Sefc et al. 1998, Cseh et al. 2006, de Andrés et al. 2007, Dzhambazova et al. 2007, Upadhyay et al. 2007, Crespan et al. 2009, Jahnke et al. 2011).

Rootstocks plays a role in the control of pests, such as phylloxera and nematodes, and in plant adaptation to the environment, such as affecting scion resistance to abiotic stresses (Vršič et al. 2015, Ollat et al. 2015). These aspects make choosing the right rootstock/scion combination very challenging. Nevertheless, the low information available for the non-*vinifera* sector makes this choice entrusted to a few elite

rootstocks. For this reason, an accurate genetic characterization of the non-*vinifera* germplasm is required.

In 2013, at the University of Milan a large collection of 379 rootstocks and other non-*vinifera* accessions was established. It is one of the largest collections in Italy, with accessions coming from other repositories in Italy, Spain and USA. Notably, this collection encompasses the progenies obtained so far from the University of Milan breeding programs and recorded as the "F series". It was established to gather as wide as possible genetic variability of those *Vitis* species used for rootstock breeding, and to identify the most suitable germplasm for the new breeding programs to adapt viticulture to climate changes (such as drought; Bianchi et al. 2018).

The aim of this work was to shed light on the identity of non-*vinifera* germplasm included in the rootstock collection of University of Milan, indicating redundancies and uniqueness for a better management of the collection itself. The 379 accessions were phenotyped for the sex of flowers and genotyped by 22 SSR loci and allelic profiles of only 17 SSR markers were used for genotype identification, to investigate the genetic diversity and parentage, to infer the structure of this population and to set up a core collection able to capture the whole allelic variation.

Materials and Methods

Plant material

A total of 379 grapevine rootstock accessions were analyzed, including wild non-*vinifera Vitis* species. They belonged to the ampelographic collection of the University of Milan (Riccagioia, Torrazza Coste, Pavia, Italy), and were collected mainly from the

following donor institutes in Italy, Spain and USA: CREA of Conegliano (Council for Agricultural Research and Agricultural Economy Analysis - Research Centre for Viticulture and Enology, Conegliano, Italy), FEM (Edmund Mach Foundation, San Michele all'Adige, Trento, Italy), IMIDRA (Instituto Madrileño de Investigación y Desarrollo Rural, Agrario y Alimentario, El Encín, Alcalá de Henares, Madrid, Spain), CIFA (Centro de Investigación y Formación Agraria, Rancho de la Merced, Jerez de la Frontera, Cádiz, Spain), and USDA (United States Department of Agriculture - Agricultural Research Service, Davis, California) (Table S1). The collection is located in the Oltrepò Pavese viticultural area (longitude, E 9° 05′; latitude, N 44° 58′), 144 m above sea level. The training system is free bush and the rows are north–south oriented. The soil is clay textured. Plants were spaced at 2.5 m (inter-row) × 1 m (intra-row).

Phenotyping

The sex of flowers of the accessions was ascertained in 2016 and 2017, in accordance with the character code 151 of the 2nd edition of the "OIV Descriptor list for grape varieties and *Vitis* species" (OIV 2009).

DNA extraction

Fifty milligrams of freeze-dried young leaf tissue were ground with a Qiagen TissueLyser system (Qiagen, Hilden, Germany) and total genomic DNA was isolated using a DNeasy 96 Plant Kit (Qiagen), according to the manufacturer's protocol. DNA quality and concentration were checked spectrophotometrically and by electrophoresis

on agarose gel. DNA samples were suspended in TE buffer (pH = 8) and diluted to approximately 10 ng/µl for PCR.

SSR selection and amplification

The genotyping of the 379 accessions was performed using 22 SSR markers: i) the international set of nine SSR loci adopted by the OIV (VVS2 (Thomas and Scott 1993); VVMD5, VVMD7, VVMD25, VVMD27, VVMD28 and VVMD32 (Bowers et al. 1996, 1999); ii) VrZAG62 and VrZAG79 (Sefc et al. 1999)); iii) VMC6E1 and VMC6G1 (Crespan 2003); iv) VMCNG4b9 (Welter et al. 2007); v) three VrZAG markers (VrZAG25, VrZAG64 and VrZAG83 (Sefc et al. 1999)); vi) seven VChr SSRs ((VChr1b, VChr3a, VChr4a, VChr7b, VChr10b, VChr13c and VChr19a (Cipriani et al. 2008)). Seven VChr markers were selected from a set of 26 markers after analysis of polymorphism information content (PIC), clarity of allelic peaks and lower probability of null alleles with a set of 26 rootstocks (Table S1 and S2). This set of 22 SSR markers was selected because: i) the nine OIV loci are used worldwide for *V. vinifera* cultivar identification; ii) the three VMC loci are routinely use by the authors for *V. vinifera* cultivar identification (Migliaro et al. 2013); iii) the three VrZAG loci are obtained from *V. riparia* (Sefc et al. 1999); iv) the seven VChr loci are suggested for an easier genotyping in *V. vinifera* (Cipriani et al. 2008) and not yet tested on non-*vinifera* varieties.

PCR reactions were performed using forward primers labeled with fluorescent dyes (6-FAM, PET, VIC, or NED); three multiplex panels of fluorescent-labeled microsatellite loci were used. Simultaneous PCR amplifications were conducted in a final volume of

12.5 µl containing 1× PCR reaction buffer, 10 ng of genomic DNA, 0.2 mM of each dNTPs, 2 mM $MgCl_2$, 1.5 U Taq DNA Polymerase (Qiagen). Depending on the locus, primer concentrations ranged from 0.11 to 0.48 µM. Reactions were performed on a GeneAmp PCR System 9700 using the following profile: a hot start of 95 °C for 5 min, 30 amplification cycles of 45 sec at 95 °C, 1 min at 55 °C, 30 sec at 72 °C, and a final extension step of 30 min at 72 °C. Presence of PCR products was assessed by electrophoresis in a 1.5% agarose gel and quantified by comparison with Mass ruler DNA ladder mix (Thermo Fischer Scientific, Waltham, MA). PCR products (0.5 µl) were mixed with 9.35 µl of formamide and 0.15 µl of the GeneScan™ 500 LIZ Size Standard (Life Tech, Foster City, CA). Capillary electrophoresis was conducted in an ABI 3130xl Genetic Analyzer (Life Tech). Allele calling was performed with GeneMapper 4.0, using the 500 LIZ size standard as internal ladder and a homemade binset built with reference varieties. Allele sizes were recorded in bp and genotypes showing a single peak at a given locus were considered as homozygous.

Statistics on SSR data

Cervus 3.0 software (Kalinowski et al. 2007), and GenAlEx 6.5 software (Peakall and Smouse 2006) were used to determine the number of different alleles (*No*), effective number of alleles (*Ne*), observed (*Ho*) and expected heterozygosity (*He*), polymorphic information content (PIC), Hardy Weinberg equilibrium (HW), the probability of null alleles (F(null)), identity probability between two unrelated individuals and between two hypothetical full siblings (P_{ID} and P_{ID}SIBS).

Genotype accumulation curve

An accumulation curve approach, obtained by rarefaction, was adopted to identify the minimum number of SSR loci to explain the observed diversity in our dataset. The function *genotype_curve* of R (R Core Team 2014) package *poppr* (Kamvar et al. 2014) was used, randomly sampling *x* loci per *r* times and counting the number of observed different genotypes. *r* value was set to 1000. The results were plotted as barplot and a trend line was added using the *ggplot2* function *stat_smooth* (Wickham 2009) in R.

Identification and pedigree relationships

True-to-types and misnomers genotypes were identified according to the information available in the literature and public database: i) the *Vitis* International Variety Catalogue (*V*IVC, http://www.vivc.de); ii) "Pl@ntGrape, le catalogue des vignes cultivées en France" (http://plantgrape.plantnet-project.org/fr); iii) the U.S. National Plant Germplasm System (NPGS, https://npgsweb.ars-grin.gov/gringlobal/search.aspx). Other identifications were made overlapping different information: i) SSR profile of accessions coming from different repositories and sharing the same name; ii) the declared parents; iii) the flower sex.

Possible PO (Parent-Offspring) relationships were obtained with GenAlEx 6.5 software. Putative GO (Genitors-Offspring) trios were produced with Cervus 3.0 (Kalinowski et al. 2007). Flower sex information supported the pedigree relationships proposed.

Population structure and core collection

The genetic structure of the germplasm collection was analyzed with STRUCTURE 2.1 software (Pritchard et al. 2000) and with the R package *adegenet* (Jombart 2008) . STRUCTURE software follows a Bayesian approach to interpret the correlation among genotypes in terms of admixture between a defined number of ancestral populations. The burn-in period consisted of 500,000 iterations followed by 100,000 MCMC (Markov Chain Monte Carlo) repeats. The number of clusters (*K*, number of ancestral genetic groups) varied from 2 to 10 and ten replicate runs were conducted to quantify the variation of the likelihood for each *K*. The most likely *K* value was chosen according to the "Evanno's method" (Evanno et al. 2005) implemented in the Structure Harvester software (Earl and vonHoldt 2012). *adegenet* performs Principal Coordinate Analysis (PCoA), based on standardized covariance of genetic distances calculated for codominant markers. The samples were arranged based on their membership to the STRUCTURE clusters. The genetic relationships among genotypes were displayed by a two-dimensional scatter plot.

In order to select a subsample of a large germplasm collection that contains the minimum number of individuals representing the whole genetic diversity of the original collection, a core collection was assessed (Brown 1995). To design the core collection, only the unique genotypes were included in the analysis. The allele coverage allocation strategy implemented in R package *corehunter* (Thachuk et al. 2009) was used to generate the genetic core collections, maximizing the proportion of observed alleles in the entire SSR dataset. Different core subsets were generated, changing the *size* parameter of the desired core collections. The sizes ranged from 0.1 to 0.9 of the entire dataset. Per each core collection, number of alleles, Nei's genetic index (Nei 1987) and Ho were determined by GenAlEx 6.5 software (Peakall and Smouse 2006).

Results

A total of 379 grapevine rootstocks and other non-*vinifera* accessions maintained in the University of Milan germplasm collection were characterized by phenotyping the sex of flowers and genotyping 22 SSR markers. Genotype identification was performed, as far as possible. A minimal set of SSR markers useful for non-*vinifera* grapevine varieties identification was computed. Genotype identification was assessed, as well as genetic diversity and population structure. A genetic core collection capturing the maximum allelic diversity was set up.

Phenotyping

Flower sex was checked on 219 out of 379 accessions, about 94% of the unique genotypes (see "Genetic diversity" section; Table S3). About 46% of genotypes were male and 44% female. The other less representative phenotypes were hermaphrodite (7.8%) and male flowers with reduced gynoecium (2.3%).

Selection of SSR loci for grapevine rootstock fingerprinting

Twenty-two SSR markers were used to genotype the entire collection. This set included nine OIV, three VMC, three VrZAG and seven VChr SSR loci. Five out of the initial 22 SSR markers were subsequently discarded (VVMD32, VMC6G1, VChr1b, VChr13c, and VrZAG25), because problems were evidenced on the large-scale analysis of the whole set of accessions, such as no amplification in some samples (VChr1b), amplification of more than two alleles (VChr13c, VrZAG25) and very low signals for

231 particular alleles that could be omitted from allele calling (VVMD32 and VMC6G1).
232 Some mono-locus in *vinifera* SSRs (VChr8a, VChr13c, VrZAG25 and VMCNG4b9),
233 clearly amplified more than one locus in non-*vinifera* genotypes, showing more than
234 two alleles per sample. More detailed technical comments are reported in the
235 Supplemental text 1.

236 Finally, 17 SSRs were retained and used for statistical data processing. They included
237 eight out of nine "international" (VVS2, VVMD5, VVMD7, VVMD25, VVMD27,
238 VVMD28, VrZAG62 and VrZAG79), two VMC (VMC6E1 and VMCNG4b9), two
239 VrZAG (VrZAG64 and VrZAG83) and five long-repeat VChr (VChr3a, VChr4a,
240 VChr7b, VChr10b and VChr19a) microsatellite markers.

241

242 Genetic diversity

243 Genotyping of 379 accessions with 17 selected SSRs produced 232 different molecular
244 profiles (Table S4). About 39% of accessions in the collection represented redundant
245 germplasm. The accessions were grouped and ordered depending on the genotype and
246 identification results are reported in Table S3.

247 Statistics on the selected 17 SSRs are shown in Table 1, while the lists of alleles
248 per locus are reported in Table S5. Three hundred and seventy-three alleles were
249 detected, with a mean of 21.94 per locus and a mean number of effective alleles of 8.52.
250 The mean Ho and mean He were very similar (0.85 and 0.86, respectively). The
251 probability of identity was 2.33×10^{-27}. Out of the 373 SSR alleles, 29.7% displayed a
252 frequency higher than 5% and were classified as 'common' alleles, 34.3% showed
253 frequencies between 1% and 5% and were 'less common' alleles, and 35.9% had a

frequency lower than 1% and were ‘rare alleles’, suggesting that this collection includes a broad biodiversity.

Genotype accumulation curve

A genotype accumulation curve was drawn to detect the minimum number of loci able to capture the whole genetic diversity in the dataset of 232 grapevine genotypes. The results are plotted in Figure S1. 99.9% of genetic diversity was reached with seven SSR loci, indicating that additional markers are unnecessary to distinguish genotypes.

The genotype identification

Obtaining the correct correspondence between molecular profile and variety name, when available, has proven to be very difficult, due to the poor information on many accessions. Even pedigree records, which could greatly support this exploration, evidenced many limits, mainly due to the following issues: a) generic information on the parents used for crossing, b) some rootstocks derived from open pollination, c) one or both parents of the bred variety were different from those declared. Sixty-nine genotypes were identified according to the information already available in the literature and databases. Another 44 genotypes were identified combining the available information, such as SSR profile of accessions coming from different repositories and sharing the same name (SSR profile ID 63 and others), the declared parents (SSR profile ID 150 and others) and the flower sex (SSR profile ID 193 and others) (References column in Table S3). For some genotypes, such as M1, M2, M3 and M4 (bred at the University of Milan) and Cosmo 4, Cosmo 7 and Cosmo 9 (bred at the

277 CREA of Conegliano, Treviso), the SSR profiles were reported for the first time (Table
278 S3, SSR profiles ID 64, 65, 66, 67 and 225, 226, 227, respectively). For the other
279 genotypes, a progressive genotype number was proposed (such as genotype 1 for the
280 SSR profile ID 68 in Table S3). For a detailed description of genotype identification,
281 please refers to the Supplementary text 2.

282

283

284 Pedigree relationships

285 Putative genitors and offspring (GO trios) relationships were obtained using Cervus
286 software. After discarding the GO trios showing obvious inconsistencies, such as sexual
287 incompatibility of the proposed genitors or historical information on the chronological
288 appearance of the varieties, 77 GO trios were proposed (Table S3). Some of them were
289 confirmed pedigrees (SSR profile ID 5 and other), some others partially confirmed
290 (SSR profile ID 25 and other), updated (SSR profile ID 7 and other), suggested (SSR
291 profile ID 53 and other) and discarded (SSR profile ID 2 and other). For a detailed
292 description of kinship results, please refers to Supplementary text 3.

293 XX parent-offspring (PO) relationships were identified (Table S3). These PO
294 relationships were mainly collected for unidentified genotypes (SSR profile ID XX and
295 other) or crosses obtained by the University of Milan (SSR profile ID XX and other).
296 The results indicated that the parents recorded during the breeding activities do not
297 always find a positive feedback (such as SSR profile ID 87 and other). These
298 preliminary indications need further validation by increasing the number of markers.

299

300 Population structure and core collection

301 In order to ascertain the likely number of genetic groups (*K*) within the data sample, a
302 structure analysis was performed and ancestral values per genotype were recorded for *K*
303 ranging from 2 to 10. Based on Evanno's method (Evanno et al. 2005), the most likely
304 number of populations was estimated at $K = 3$, showing the largest value in the ΔK
305 (19.792). The three clusters are shown in the barplot in Figure 1. Setting at $Q > 0.85$ the
306 threshold for group assignment, 147 genotypes (around 63%) were univocally assigned
307 to a cluster at $K = 3$. The proportion of admixed genotypes resulted around 37% (Table
308 S6). The three refined groups accounted for 40, 31 and 29% of the genotypes univocally
309 assigned to a population. Cluster 1 encompassed a highly heterogeneous mix of *Vitis*
310 *rupestris* and *V. riparia* varieties. The genotypes of Cluster 2 were represented by the
311 three main varieties used worldwide for breeding, i.e. Rességuier n. 2, Gloire de
312 Montpellier, Rupestris du Lot and their progenies. Cluster 3 included mostly *V.*
313 *labrusca* hybrids, such as Isabella and Jacquez, and hybrids with *V. vinifera*, such as 41
314 B, Fercal, Golia, Blanchard 1 B, Berlandieri-Colombard 2. As expected, the largest
315 percentage of individuals with a complex pedigree (where more than two *Vitis* genomes
316 can be accounted for) was not successfully assigned to a specific ancestral population
317 and were classified as admixed.

318 The results of PCoA are graphically reported in Figure 2. The genotypes are
319 represented by four different colors, corresponding to the four groups (three ancestral
320 populations plus one admixed group) recorded by structure analysis and listed in Table
321 S6. The first two principal coordinates accounted for 15.66% of total variability, due to
322 the high percentage of admixed genotypes and the use of "elite" varieties for breeding.
323 Even though there was overlapping among the clusters (Cluster 3 and Cluster 4

remained in between Cluster 1 and Cluster 2), Cluster 1 and Cluster 2 were shown to be very well discriminated also by PCoA.

Nine different core collections were built up with a collection size ranging from 90 to 10% of the entire dataset to identify the smallest set of genotypes as representative as possible of available genetic diversity (Table 2). Seventy genotypes (Core3) were enough to capture 100% of the 373 detected alleles, while decreasing the core collection size up to 0.1, the number of alleles captured were 299 (about 20% less than the total alleles detected). Nei's genetic index ranged from 0.531 (Core1) to 0.962 (Core9), with the Core3 reaching 0.733. The Ho values were quite similar to that detected for the entire collection (Core10 = 0.847), varying from 0.808 (Core1) to 0.864 (Core9). Core3 showed the value 0.823. Table S7 lists the genotypes included in the nine different core collections. Referring to the refined clusters in Table S6, the varieties selected in Core3 were: 37.1 % from the admixed group, 30.0 % from Cluster 3, having *labrusca* and *vinifera* parentage, 25.7 % from Cluster 1, which are non-*vinifera* varieties, and only 7.1% from Cluster 2, the so-called "Berlandieri x Rupestris" and "Berlandieri x Riparia" varieties.

Discussion

Three hundred and seventy-nine accessions of the University of Milan rootstock collection were phenotyped for the sex of flowers and genotyped based on 22 SSR markers. Even excluding the most recent crosses, the difficulty in identification emerged from the notable number of genotypes remained anonymous and the misnomered accessions.

Phenotyping results showed the typical dioecious trait of non-*vinifera* grapevine germplasm, with about 90% of genotypes divided between male and female plants. Most of the hermaphrodite genotypes were shown to carry a *V. vinifera* parentage, such as Berlandieri-Colombard 2 (SSR profile ID 27), Golia (SSR profile ID 34), 1132 Grimaldi (SSR profile ID 182). On the other hand, this trait can be explained in some unidentified genotypes by the possible PO relationship with rootstocks having *V. vinifera* parentage, like Ganzin 1.

Microsatellites are a valuable tool for the cultivar identification. They were largely used to successfully genotype the *vinifera* germplasm (Laucou et al. 2011, Cipriani et al. 2008, Emanuelli et al. 2013, Riaz et al. 2018). For the *vinifera* compartment, a set of nine SSR loci has been proposed as universal set for cultivar identification (Maul et al. 2012). Regarding non-*vinifera* compartment, a lot of information is still missing. Therefore, in this paper great attention was made to identify the most suitable set of SSR loci to genotype the non-*vinifera* individuals and to identify the minimum number of molecular markers to be used for cultivar identification. Up to 22 SSR loci were chosen to analyze the entire rootstock collection, but only 17 loci (eight OIV, two VMC, two VrZAG and five VChr loci) were suitable for genotyping. Recently, SSR loci with core repeats three to five nucleotides long were proposed for *vinifera* cultivar identification, because neighbor alleles are more easily distinguished from each other (Cipriani et al. 2008, 2010). After a preliminary selection of seven VChr markers out of the 26 initially tested, we were able to propose a set of five long core-repeat SSRs useful for non-*vinifera* variety genotyping. The genotype accumulation curve analysis showed that randomly sampling seven out of 17 loci analysed all the non-*vinifera*

370 genotype are distinguished. Discarding VVMD25 and VVMD32, the remaining seven
371 "international" SSR loci can be used as the minimal set for rootstock identification.

372

373 s. .

374 .

375 Genotyping the rootstock collection with 17 SSRs, 232 molecular profiles were
376 identified (Table S3 and Table S4), showing a rich genetic diversity. Three hundred and
377 seventy-three alleles were recorded, having a high percentage of less common and rare
378 alleles (70.2%). This result explained why only 7 random SSRs were enough to
379 distinguish all 232 recorded genotypes. On average, we found 21.9 alleles per locus and
380 a mean Ne of 8.52, high values despite the low number of genotypes. Our data matched
381 the results obtained by Laucou et al. (2011) on 135 rootstocks and Emanuelli et al.
382 (2013) on 127 rootstock genotypes analysed with 20 and 22 SSRs, respectively: they
383 found 405 and 412 alleles in total, 20.25 and 18.7 alleles per locus on average.
384 Emanuelli et al. (2013) computed a Ne mean value of 8.199, which is very similar to
385 our result (8.52). Instead, we computed a higher Ho mean value, 0.84 *versus*0.78 and
386 0.73 computed by Laucou et al. (2011) and Emanuelli et al. (2013), respectively. Both
387 groups of authors underlined that genetic diversity of non-*vinifera* varieties was higher
388 than the domesticated sector, as can be expected by the heterogeneity of rootstock
389 genotypes.

390 Literature and database allowed 29.7% of genotypes to be identified, and we were able
391 to propose additional assignments (44) by combining different information. Cultivars
392 represented here by single accessions could include mistakes and many genotypes

393 remained anonymous (34.4%, without taking into account recent, still unnamed
394 crosses). Previous studies and the present one showed clear problems of correct
395 identification in rootstock repositories. These problems are related to mislabeling,
396 multiple naming of single cultivars, and incorrect denominations. Molecular analyses
397 can support identification, but these results need to be integrated with ampelographic
398 observations. In this study, the sex of flowers has been used as one useful
399 ampelographic trait for this goal. In the past, many breeders worked hardly producing
400 an enormous quantity of crosses, often with poorly controlled management. This left us
401 a complex heritage and we now find it difficult to ascertain the correct identity of many
402 materials. The declared pedigrees are not necessarily a reliable tool, either because they
403 are often too generic (such as Riparia x Rupestris) or there is no match between the
404 declared parents and the true ones. Even if 17 markers were not enough to draw out
405 robust conclusions on first degree relationships, they were useful to support or question
406 known information, and to suggest new possible parentages. The large amount of SSR
407 profiles produced in this study contributed to advance the identification of rootstocks.
408 This information can also be useful for future applications, such as the study of pedigree
409 relationships to better address the choice of resistance donors for conventional breeding
410 programs (Di Gaspero et al. 2012).

411 The large number of rootstocks available for viticulture is the result of many
412 interspecific crosses among vines belonging mainly to *V. berlandieri, V. riparia* and *V.*
413 *rupestris* species. The genetic structure of grapevine rootstocks has been strongly
414 influenced by human selection during the post-phylloxera breeding programs. Structure
415 analysis identified three significant genetic groups, without a clear discrimination based
416 on the species. A large majority of genotypes remained admixed, evidencing the genetic

417 complexity of the plant material analyzed. PCoA clearly separated Gloire de
418 Montpellier, Rességuier n. 2 and Rupestris du Lot, the three main founders of the
419 currently most widespread rootstocks and related progenies, from all the other
420 genotypes. These three founders were from three different *Vitis* species: *V. riparia*, *V.*
421 *berlandieri* and *V. rupestris*. Nevertheless, being just three varieties, the genetic
422 variability of the progenies, which had always reorganized the same genotypes, is low
423 compared to the other genetic materials. This lower variability was underlined by
424 establishing the core collection. The core collections represent the ability to capture the
425 diversity of the entire dataset in a restricted set of genotypes that is feasible to handle.
426 Different core collections have been developed for the *V. vinifera* cultivated sector, such
427 as those including genotypes from the Vassal collection (Le Cunff et al. 2008), and
428 from Italian collections (Cipriani et al. 2010, Emanuelli et al. 2013). In this study, we
429 sorted out a first core collection related to grapevine rootstocks. Seventy out of 232
430 genotypes were included in the most suitable core collection, capturing the entire allelic
431 richness. Interestingly, the highest number of genotypes (26) was kept from those
432 classified as admixed after structure analysis, then from Cluster 3 (21) and Cluster 1
433 (18). Only five varieties were from Cluster 2, because this cluster combined the lowest
434 genetic variability with the highest redundancy of over represented genotypes.

435 Among the 70 genotypes of Core3 only 29 were identified varieties. Ampelographic and
436 agronomic characteristics of all the other genotypes are still lacking. Nevertheless, a few
437 of 29 identified varieties were grapevine rootstocks commonly used in viticulture, and
438 only Riparia Gloire de Montpellier was included from the three main progenitors. The
439 purpose of building up a core collection was to selected entries to be further investigated
440 for their use in breeding programs for a sustainable Mediterranean viticulture. Based on

441 our core collection, the new breeding programs could rely on genotypes with a
442 phenotypic diversity not yet exploited. Additionally, this genetic core collection could
443 be a valuable starting point for future phenotyping efforts and genome-wide association
444 studies.

445 Despite the presence of large kin groups in the dataset, the number of genotypes
446 required to sort out the core collection covering the totality of alleles found (373) was
447 high (about 30%). This result can be related to the amount of varieties belonging to
448 different *Vitis* species and the high genetic variability detected in our dataset (high mean
449 number of alleles per locus in comparison to the low number of genotypes). In *V.*
450 *vinifera* subsp. sativa core collections the same result was obtained, retaining 4% of
451 genotypes in Le Cunff et al. (2008), 4.7% in Cipriani et al. (2010), and 15% in
452 Emanuelli et al. (2013).

453 Reynolds and Wardle (2001) outlined seven major criteria for choosing rootstocks: i)
454 phylloxera resistance; ii) nematode resistance; iii) adaptability to high pH soils; iv)
455 adaptability to saline soils; v) adaptability to low pH soils; vi) adaptability to wet or
456 poorly drained soils; vii)adaptability to drought. Other important requirements are
457 rooting ability, grafting affinity and biomass production. Numerous reports have also
458 proved that rootstocks affect vine growth, yield and fruit and wine quality (Boselli et al.
459 1992, Walker et al. 1998, 2000, Gregory et al. 2013, Berdeja et al. 2015). These effects
460 are direct and are the consequences of the interactions between environmental factors,
461 scion physiology and rootstock cultivar used. Selecting the suitable rootstock for a
462 vineyard is just as important as variety selection. The results presented in this paper
463 confirm that the rootstocks used world-wide today mostly derive from a very narrow
464 genetic base. Additional problems, including the appearance of new phylloxera

biotypes, are serious threats since present rootstocks may not provide the genetic diversity necessary to counteract evolving soil-borne pests (Pap et al. 2018). Future association studies and marker-assisted selection for resistance to phylloxera, drought, salt, and other traits, can gain advantage from the information provided in this paper, because it contributes to: i) share the number of different genotypes available in the repositories with the scientific community; ii) define their true-to-type; iii) ascertain their origin and genetic diversity; iv) correctly associate present and future phenotypes to the right genotypes. The new achievements will lengthen the shortlist of rootstocks used today to keep pace with the development of sustainable viticulture models.

Conclusion

This paper contributed to ascertain how many genotypes were growing in the large rootstock collection of the University of Milan, to identify them as a starting point for the exploitation of still unexplored genetic resources. The large number of unidentified genotypes, the population structure and the core collection show the potential of this repository to hold individuals useful for new breeding programs. Genotypes with a phenotypic diversity not exploited yet can become the starting germplasm to obtain new rootstocks able to cope with some factors crucial for developing sustainable viticulture in the Mediterranean area. A phylloxera risk limitation can be implemented by avoiding the use of rootstocks carrying *vinifera* background in their pedigree.

Since low information is available regarding the most suitable SSR molecular markers to be used for the genetic analysis of non-*vinifera* species, this work can be considered pioneer for this purpose.

Literature Cited

Berdeja M, Nicolas P, Kappel C, Dai ZW, Hilbert G, Peccoux A, Lafontaine M, Ollat N, Gomès E and Delrot S. 2015. Water limitation and rootstock genotype interact to alter grape berry metabolism through transcriptome reprogramming. Hortic Res 2: 15012.

Bianchi D, Grossi D, Tincani DTG, Simone Di Lorenzo G, Brancadoro L and Rustioni L. 2018. Multi-parameter characterization of water stress tolerance in *Vitis* hybrids for new rootstock selection. Plant Physiol Biochem 132:333-340.

Boselli M, Fregoni M, Vercesi A and Volpe B. 1992. Variation in mineral composition and effects on the growth and yield of Chardonnay grapes on various rootstocks. Agricoltura Ricerca 14:138-139.

Bowers JE, Dangl GS, Vignani R and Meredith CP. 1996. Isolation and characterization of new polymorphic simple sequence repeat loci in grape (*Vitis vinifera* L.). Genome 39:628-633.

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

510 Brown AHD. 1995. The Core Collection at the Crossroads. Hodgkin T, Brown AHD,
511 van Hintum TJL and Morales EAV, eds. Core Collections of Plant Genetic Resources.
512 (John Wiley and Sons, Hoboken) pp. 3-19.

513

514 Cipriani G, Marrazzo MT, Di Gaspero G, Pfeiffer A, Morgante M and Testolin R. 2008.
515 A set of microsatellite markers with long core repeat optimized for grape (*Vitis* spp.)
516 genotyping. BMC Plant Biol 8:127.

517

518 Cipriani G, Spadotto A, Jurman I, Di Gaspero G, Crespan M, Meneghetti S, Frare E,
519 Vignani R, Cresti M, Morgante M, Pezzotti M, Pe E, Policriti A and Testolin R. 2010.
520 The SSR-based molecular profile of 1005 grapevine (*Vitis vinifera* L.) accessions
521 uncovers new synonymy and parentages, and reveals a large admixture amongst
522 varieties of different geographic origin. Theor Appl Genet 121:1569-1585.

523

524 Cosmo I. 1958. Portinnesti della vite. Edizioni Agricole, Bologna, Italy.

525

526 Crespan M. 2003. The parentage of Muscat of Hamburg. Vitis 42:193-197.

527

528 Crespan M, Meneghetti S and Cancellier S. 2009. Identification and genetic relationship
529 of the principal rootstocks cultivated in Italy. Am J Enol Vitic 60:349-356.

530

531 Cseh A, Taller J, Podmaniczky P, Kocsis L. 2006. Comparative analysis of the most
532 widespread grapevine rootstock lines in the world, the Teleki lines, with microsatellite
533 markers. Cereal Res Commun 34(1):773-776.

534

535 de Andrés MT, Cabezas JA, Cervera MT, Borrego J, Martinez-Zapater JM and Jouve N.
536 2007. Molecular characterization of grapevine rootstocks maintained in germplasm
537 collections. Am J Enol Vitic 58:75-86.

538

539 Di Gaspero G, Copetti D, Coleman C, Castellarin SD, Eibach R, Kozma P, Lacombe T,
540 Gambetta G, Zvyagin A, Cindrić P, Kovács L, Morgante M and Testolin R. 2012.
541 Selective sweep at the *Rpv3* locus during grapevine breeding for downy mildew
542 resistance. Theor Appl Genet 124:277-286.

543

544 Dzhambazova T, Hvarleva T, Hadjinicoli A, Tsvetkov I, Atanassov A and Atanassov I.
545 2007. Characterization of grapevine rootstocks using microsatellite markers. Biotechnol
546 Biotechnol Equip 21:58-62.

547

548 Earl DA and VonHoldt BM. 2012. Structure Harvester: a website and program for
549 visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet
550 Resour 4:359-361.

551

552 Emanuelli F, Lorenzi S, Grzeskowiak L, Catalano V, Stefanini M, Troggio M, Myles S,
553 Martinez-Zapater JM, Zyprian E, Moreira FM and Grando MS. 2013. Genetic diversity
554 and population structure assessed by SSR and SNP markers in a large germplasm
555 collection of grape. BMC Plant Biol 13:39.

556

557 Evanno G, Regnaut S and Goudet J. 2005. Detecting the number of clusters of
558 individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611-
559 2620.

560

561 Galet P. 1988. Cépages Et Vignobles de France - Tome 1: Les Vignes Américaines.
562 Imprimerie Dehan, Montpellier, France.

563

564 Garris A, Cousins P, Ramming D and Baldo A. 2009. Parentage analysis of Freedom
565 rootstock. Am J Enol Vitic 60:357-361.

566

567 Granett J, Walker MA, Kocsis L and Omer AD. 2001. Biology and management of
568 grape Phylloxera. Annu Rev Entomol 46:387-412.

569

570 Gregory PJ, Atkinson CJ, Bengough AG, Else MA, Fernández-Fernández F, Harrison
571 RJ and Schmidt S. 2013. Contributions of roots and rootstocks to sustainable,
572 intensified crop production. J Exp Bot 64:1209-1222.

573

574 Jahnke G, Kocsisné Molnár G, Májer J, Szöke B, Tarczal E, Varga P and Kocsis L.
575 2011. Analysis of grape rootstocks by SSR markers. J Int Sci Vigne Vin 45:199-210.

576

577 Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers.
578 Bioinformatics 24:1403-1405.

579

580 Kalinowski ST, Taper ML and Marshall TC. 2007. Revising how the computer program
581 CERVUS accommodates genotyping error increases success in paternity assignment.
582 Mol Ecol 16:1099-1106.

583

584 Kamvar ZN, Tabima JF and Grünwald NJ. 2014. Poppr: an R package for genetic
585 analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ
586 2:e281.

587

588 Laucou V, Boursiquot JM, Lacombe T, Bordenave L, Decroocq S and Ollat N. 2008.
589 Parentage of grapevine rootstock 'Fercal' finally elucidated. Vitis 47:163-167.

590

591 Laucou V, Lacombe T, Dechesne F, Siret R, Bruno JP, Dessup M, Dessup T, Ortigosa
592 P, Parra P, Roux C, Santoni S, Vares D, Peros JP, Boursiquot JM and This P. 2011.
593 High throughput analysis of grape genetic diversity as a tool for germplasm collection
594 management. Theor Appl Genet 122:1233-1245.

595

596 Le Cunff L, Fournier-Level A, Laucou V, Vezzulli S, Lacombe T, Adam-Blondon AF,
597 Boursiquot JM and This P. 2008. Construction of nested genetic core collections to
598 optimize the exploitation of natural diversity in *V. vinifera* L. subsp sativa. BMC Plant
599 Biol 8:31.

600

601 Lin H and Walker MA. 1998. Identifying grape rootstocks with Simple Sequence
602 Repeat (SSR) DNA markers. Am J Enol Vitic 49:403-407.

603

604 Maul E, Sudharma KN, Kecke S, Marx G, Müller C, Audeguin L, Boselli M,
605 Boursiquot JM, Bucchetti B, Cabello F, Carraro R, Crespan M, de Andrés MT, Eiras
606 Dias J, Ekhvaia J, Gaforio L, Gardiman M, Grando S, Gyropoulos D, Jandurova O, Kiss
607 E, Kontic J, Kozma P, Lacombe T, Laucou V, Legrand D, Maghradze D, Marinoni D,
608 Maletic E, Moreira F, Muñoz-Organero G, Nakhutsrishvili G, Pejic I, Peterlunger E,
609 Pitsoli D, Pospisilova D, Preiner D, Raimondi S, Regner F, Savin G, Savvides S,
610 Schneider A, Sereno C and Simon S. 2012. The European *Vitis* Database (www.eu-
611 vitis.de) – a technical innovation through an online uploading and interactive
612 modification system. Vitis 51:79-86.

613

614 Migliaro D, Morreale G, Gardiman M, Landolfo S and Crespan M. 2013. Direct
615 multiplex PCR for grapevine genotyping and varietal identification. Plant Genet Resour
616 11:182-185.

617

618 Nei M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York,
619 USA.

620

621 Ollat N, Peccoux A, Papura D, Esmenjaud D, Marguerit E, Tandonnet JP, Bordenave L,
622 Cookson SJ, Barrieu F, Rossdeutsch L, Lecourt JL, Lauvergeat V, Vivin P, Bert PF and
623 Delrot S. 2015. Rootstocks as a component of adaptation to environment. Gerós H,
624 Chaves MM, Medrano Gil H and Delrot S. eds. Grapevine in a Changing Environment:
625 A Molecular and Ecophysiological Perspective. Wiley-Blackwell pp. 68-108.

626

627 Organisation Internationale de la Vigne et du Vin (OIV). 2009. OIV Descriptor List for
628 Grape Varieties and Vitis Species. 2d ed. Organisation Internationale de la Vigne et du
629 Vin: Paris, France. http://www.oiv.int/oiv/info/frplubicationoiv.

630

631 Pap D, Riaz S, Walker A. 2018. Living on the edge: the narrow genetic base of the
632 rootstsocks is a serious threat. Abstract book for the XII International Conference on
633 Grapevine Breeding and Genetics, Bordeaux (France), pag. 104.

634

635 Peakall R and Smouse PE. 2006. GenAlEx 6: genetic analysis in Excel Population
636 genetic software for teaching and research. Mol Ecol Notes 6:288-295.

637

638 Poczai P, Hyvönen J, Taller J, Jahnke G and Kocsis L. 2013. Phylogenetic analyses of
639 Teleki grapevine rootstocks using three chloroplast DNA markers. Plant Mol Biol
640 Report 31:371-386.

641

642 Pritchard JK, Stephens M and Donnelly P. 2000. Inference of population structure using
643 multilocus genotype data. Genetics 155:945-959.

644

645 R Core Team. 2014. R: A language and environment for statistical computing. R
646 Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

647

648 Reynolds AG and Wardle DA. 2001. Evaluation of minimal pruning upon vine
649 performance and berry composition of Chancellor. Am J Enol Vitic 52:45-48.

650

651 Santos S, Oliveira M, Amorim A and van Asch B. 2014. A forensic perspective on the
652 genetic identification of grapevine (*Vitis vinifera* L.) varieties using STR markers.
653 Electrophoresis 35:3201-3207.

654

655 Sefc KM, Regner F, Glössl J and Steinkellner H. 1998. Genotyping of grapevine and
656 rootstock cultivars using microsatellite markers. Vitis 37:15-20.

657

658 Sefc KM, Regner F, Turetschek E, Glössl J and Steinkellner H. 1999. Identification of
659 microsatellite sequences in *Vitis riparia* and their applicability for genotyping of
660 different *Vitis* species. Genome 42:367-73.

661

662 Snyder E. 1937. Grape development and improvement. USDA Yearbook of
663 Agriculture, pp 631-664.

664

665 Thachuk C, Crossa J, Franco J, Dreisigacker S, Warburton M and Davenport GF. 2009.
666 Core Hunter: an algorithm for sampling genetic resources based on multiple genetics
667 measures. BMC Bioinformatics 10:243.

668

669 Thomas M and Scott NS. 1993. Microsatellite repeats in grapevine reveal DNA
670 polymorphisms when analysed as sequence-tagged sites (STSs). Theor Appl Genet
671 86:985-990.

672

673 Upadhyay A, Saboji MD, Reddy S, Deokar K and Karibasappa GS. 2007. AFLP and
674 SSR marker analysis of grape rootstocks in Indian grape germplasm. Sci Hortic
675 112:176-183.

676

677 Vršič S, Pulko B and Kocsis L. 2015. Factors influencing grafting success and
678 compatibility of grape rootstocks. Sci Hortic 18:168-173.

679

680 Walker RR, Clingeleffer PR, Kerridge GH, Ruhl EH, Nicholas PR and Blackmore DH.
681 1998. Effects of the rootstock Ramsey (*Vitis champini*) on ion and organic acid
682 composition of grapes and wine, and on wine spectral characteristics. Aust J Grape
683 Wine Res 4:100-110.

684

685 Walker RR, Read PE and Blackmore DH. 2000. Rootstock and salinity effects on rates
686 of berry maturation, ion accumulation and colour development in Shiraz grapes. Aust J
687 Grape Wine Res 6:227-239.

688

689 Welter LJ, Göktürk-Baydar N, Akkurt M, Maul E, Eibach R, Töpfer R and Zyprian EM.
690 2007. Genetic mapping and localization of quantitative trait loci affecting fungal disease
691 resistance and leaf morphology in grapevine (*Vitis vinifera* L.). Mol Breed 20:359-374.

692

693 Wickham H. 2009. Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New
694 York, USA.

695

696 Zavaglia C, Pecile M, Gardiman M, Bavaresco L. 2016. Production of propagating
697 material of grapevine rootstocks in the EU and in Italy. Acta Hortic 1136:57-62.

698

699

Tables

Table 1 - Statistics on 17 SSR markers obtained from 232 grapevine rootstock genotypes and ordered by PIC value.

Table 2 - SSR diversity within each sample of core collections compared with the entire dataset (Core10). The most suitable core collection is highlighted in grey.

Table S1 - Preliminary results obtained using 22 VChr SSRs on 26 rootstock cultivars

Table S2 - Statistics on twenty-two VChr SSRs ranked by descending PIC values

Table S3 - List of 379 rootstock accessions genotyped with 17 SSR loci. The accessions are grouped based on their genotype. Ascertained misnomered accessions are written in italics. Flower sex and donor institutes of the accessions are reported, as well as variety name as resulted from genotyping, pedigree information from the literature and present molecular data.

Table S4 - List of alleles found at each of the 17 SSR loci analyzed

Table S5 - Allelic profiles at 17 SSR loci of 232 grapevine rootstock genotypes. Original questionable accession names are converted to "no name". In bold suggested names for uncertain varieties

Table S6 - Genotypes subdivision depending on cluster membership (Q>0,85) and admixed (Q<0.85) according to Structure. Genotypes are in alphabetical order

Table S7 - List of the genotypes included in nine different core collections, obtained from a dataset of 232 grapevine rootstock genotypes

722

723 Figure legends

724 **Figure 1**. Population structure as estimated using STRUCTURE with *K* = 3. Each
725 genotype is represented by a vertical bar, showing assignment probabilities to each of
726 the three ancestral groups.

727 **Figure 2**. Population structure based on PCoA analysis. The genotypes are arranged
728 according to the ancestral clusters identified by STRUCTURE analysis, using a Q> 0.85
729 threshold for group assignation. Cluster 1 encompasses a mix of *Vitis rupestris* and *V.*
730 *riparia* varieties; Cluster 2 groups the three main varieties used for breeding and their
731 progenies; Cluster 3 includes mostly *V. labrusca* and *V. vinifera* hybrids; Cluster 4
732 shows the admixed genotypes.

733

734 **Figure S1**. Genotype accumulation curve of 232 grapevine rootstock accessions,
735 genotyped over 16 SSR loci. The red dashed line represents 100% of the total observed
736 genotypes. Blue: trend line.

737