Theor Appl Genet (2010) 121:1569–1585
DOI 10.1007/s00122-010-1411-9

ORIGINAL PAPER

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

Guido Cipriani · Alessandro Spadotto · Irena Jurman · Gabriele Di Gaspero · Manna Crespan · Stefano Meneghetti · Enrica Frare · Rita Vignani · Mauro Cresti · Michele Morgante · Mario Pezzotti · Enrico Pe · Alberto Policriti · Raffaele Testolin

Received: 2 March 2010 / Accepted: 17 July 2010 / Published online: 6 August 2010

Abstract A collection of 1005 grapevine accessions was genotyped at 34 microsatellite loci (SSR) with the aim of analysing genetic diversity and exploring parentages. The comparison of molecular profiles revealed 200 groups of synonymy. The removal of perfect synonyms reduced the database to 745 unique genotypes, on which population genetic parameters were calculated. The analysis of kinship uncovered 74 complete pedigrees, with both parents identified. Many of these parentages were not previously known and are of considerable historical interest, e.g. Chenin blanc (Sauvignon × Traminer rot), Covè (Harslevelu selfed), Incrocio Manzoni 2–14 and 2–15 (Cabernet franc × Prosecco), Lagrein (Schiava gentile × Teroldego), Malvasia nera of Bolzano (Perera × Schiava gentile), Manzoni moscato (Raboso veronese × Moscato d'Amburgo), Moscato violetto (Moscato bianco × Duraguzza), Muscat of Alexandria (Muscat blanc à petit grain × Axina de tres bias) and others. Statistical robustness of unexpected pedigrees was reinforced with the analysis of an additional 7–30 SSRs. Grouping the accessions by profile resulted in a weak correlation with their geographical origin and/or current area of cultivation, revealing a large admixture of local varieties with those most widely cultivated, as a result of ancient commerce and population flow. The SSRs with tri- to penta-nucleotide repeats adopted for the present study showed a great capacity for discriminating amongst accessions, with probabilities of identity by chance as low as 1.45×10^{-27} and 9.35×10^{-12} for unrelated and full sib individuals, respectively. A database of allele frequencies and SSR profiles of 32 reference cultivars are provided.

Communicated by M. Frisch.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1411-9) contains supplementary material, which is available to authorized users.

G. Cipriani · G. Di Gaspero · M. Morgante · R. Testolin (✉)
Dipartimento di Scienze Agrarie e Ambientali,
University of Udine, via delle Scienze 208, 33100 Udine, Italy
e-mail: raffaele.testolin@uniud.it

A. Spadotto · I. Jurman · G. Di Gaspero · M. Morgante · A. Policriti · R. Testolin
Istituto di Genomica Applicata, Parco Scientifico e Tecnologico Luigi Danieli, via Jacopo Linussio 51, 33100 Udine, Italy

M. Crespan · S. Meneghetti · E. Frare
CRA-VIT, Centro di Ricerca per la Viticoltura,
Viale XXVIII Aprile 26, 31015 Conegliano (TV), Italy

R. Vignani · M. Cresti
Dipartimento di Scienze Ambientali,
University of Siena, via P.A. Mattioli 4, 53100 Siena, Italy

M. Pezzotti
Dipartimento di Biotecnologie, University of Verona,
Strada Le Grazie 15, 37134 Verona, Italy

E. Pe
Scuola Superiore S. Anna,
Piazza Martiri della Libertà 33, 56127 Pisa, Italy

A. Policriti
Dipartimento di Matematica e Informatica (DIMI),
University of Udine, via delle Scienze 208, 33100 Udine, Italy

Introduction

Grapevine (*Vitis vinifera* L.) is an aggregate of cultivated and feral forms distributed primarily from western Asia to Europe (Zohary and Hopf 2000). Grape domestication took place in the Neolithic, approximately 8,000 years ago, when human populations began collecting and propagating hermaphroditic and/or parthenocarpic forms, selected amongst the wild progenitor species *V. sylvestris* (Olmo 1995; This et al. 2006).

From the area of first domestication, most likely at the southern shores of the Black and Caspian seas and nearby, the domesticated forms spread westward and arrived at the Mediterranean basin following major civilisation and colonisation events (Zohary and Hopf 2000). Grape cultivation then expanded inland, reaching many temperate-cold regions of Europe and Asia (This et al. 2006).

An inventory attempted by Alleweldt in the mid-1980s revealed the existence of more than 14,000 accessions, most of which were represented by *V. vinifiera* varieties (Alleweldt et al. 1990). The origin of most of them is completely obscure because of the frequent exchange of plant material amongst the numerous secondary centres of domestication and the probable hybridisation of domesticated forms of *V. vinifera* var. *sativa* and its wild relative *V. vinifera* var. *sylvestris*, whose geographic distributions overlap (Arroyo-García et al. 2006). Until the nineteenth century, vineyards were constituted with many different varieties, and this likely gave rise to a naturally extensive hybridisation. Dozens of relatives of Pinot, Gouais and their offspring Chardonnay were reported (Bowers et al. 1999; Boursiquot et al. 2004) and many ancient superior genotypes, regarded for their quality, gave rise to large families of cultivars, such as Muscat blanc à petits grains and Muscat of Alexandria (Crespan and Milani 2001), the Malvasias (Lacombe et al. 2007), the Sangiovese (Di Vecchi Staraz et al. 2007) and others. Moreover, the extensive vegetative propagation of several ancient cultivars, such as Afus Ali, Chasselas, Grenache, Muscat, Pinot noir, Sangiovese, Sultanine and others, also gave rise to many clonal variants that contributed to the differentiation of the grapevine gene pool (Calò et al. 2001; This et al. 2006; Lacombe et al. 2007, Di Vecchi Staraz et al. 2007; Ibáñez et al. 2009).

In the second half of the nineteenth century, European vineyards were devastated by mildew diseases and phylloxera introduced from America, which caused a dramatic reduction in genetic diversity. In the twentieth century, global development of the wine grape industry further restricted the varieties in cultivation and led to the wide diffusion of a small number of French cultivars (Alleweldt et al. 1990). Many local varieties traditionally grown were abandoned in favour of varieties more adapted to the wine market demand and they have only recently been introduced back into cultivation, in homage to cultural heritage and local traditions, and to locally diversify the market.

There is a growing interest in understanding the origin and genetic diversity of germplasm rescued in different geographical areas, as well as resolving the intricacy of relationships amongst that germplasm and the most widely known international varieties.

The advent of molecular markers offered a powerful tool to address these issues, and they were largely used by ampelographers and grape geneticists. Amongst the many classes of molecular markers proposed in the last 20 years (see Schlötterer 2004 for a review), SSR or simple sequence repeat markers are the most prevalently utilised markers for genotyping individuals to solve problems of homonymy, synonymy, kinship and to infer the genetic structure of populations.

Several sets of SSR markers have been proposed in grape, the best known being the set suggested by the European group working within the grape GENRES projects (This et al. 2004). This set is based on six highly reproducible microsatellites with di-nucleotide repeats (This et al. 2004). Although the list of markers has been extended to 20 (Di Vecchi Staraz et al. 2007) and the proposed protocol appears sufficiently reliable and reproducible considering the use of reference genotypes for allele sizing, di-nucleotide repeats remain problematic due to stuttering and a narrow distance between adjacent alleles that complicates binning (Idury and Cardon 1997; Amos et al. 2007). For this reason, they have been discarded in human fingerprinting in favour of microsatellites with longer core repeats (Butler 2006).

Microsatellites with a longer core repeat are less frequent in the genome and difficult to isolate through enriched libraries, but in 2007 the complete grape genome sequence became available (Jaillon et al. 2007) and thousands of SSRs with 3-nt or longer core repeat size could be retrieved from the NCBI genebank (Cipriani et al. 2009). In a previous study, 38 SSRs with a 3- to 5-nt long core repeat were analysed and selected, and proposed as a new set for grape fingerprinting (Cipriani et al. 2008).

In the present paper, we report the analysis of 1005 grapevine accessions, which include varieties of the Italian National Grapevine Catalogue and less known local germplasm. The analysis was carried out with the new set of SSR markers with the aim of (a) verifying their informative capacity and effectiveness in discriminating amongst grape varieties, (b) creating a database of DNA profiles and allele frequencies, (c) analysing the genetic diversity existing in this collection and (d) exploring genetic kinships.

Materials and methods

Plant material

A set of 1148 grapevine (*V. vinifera* L.) accessions was collected from the national repositories of the CRA-VIT of Conegliano, TV, Italy (http://www.inea.it/isv/collezioni.html) in the locations of Susegana (latitude 45°51′09″N, longitude 12°15′27″E) and Spresiano (latitude 45°46′56″N, longitude 12°15′20″E). These accessions are representative of international, national and local varieties, half of which are registered in the Italian Catalogue of Varieties admitted to cultivation. Wild species, interspecific hybrids and rootstocks were excluded from the sample, as well as several local synonyms already known from literature.

DNA extraction and sample set reduction

DNA was extracted using the Qiagen DNeasy 96 Plant kit, and tungsten beads were used to facilitate tissue lysis. After a preliminary screening based on DNA quality, the number of samples was reduced to 1005. The list of these accessions is provided in Supplementary material S1.

The choice of SSR markers

SSR markers were selected from the list reported in Cipriani et al. (2008). Of the 38 markers, 34 described in the original paper were considered suitable for this analysis; 4 were discarded because of their low level of polymorphism or other shortcomings identified during the preliminary amplification tests.

PCR amplification and electrophoresis of SSR markers

The PCR mix was prepared in 10 µL buffer II (Applied Biosystems) using 200 µM of each dNTP, 0.2 µM of each primer, 5 ng genomic DNA and 0.5 U of Taq Gold (Applied Biosystems). The forward primer was labelled with 6-FAM, VIC, PET or NED fluorescent dye. PCR reactions were carried out in GeneAMP 9700 (Applied Biosystems) with the following thermal profile: one cycle at 95°C for 2 min, followed by 10 touchdown cycles at 94°C for 20 s, 55–0.5°C/cycle for 20 s and 65°C for 40 s, followed by 15 cycles at 94°C for 20 s, 50°C for 20 s, 65°C for 40 s and a final step of 1 h at 72°C.

PCR products were diluted in 30–60 µl H_2O, then 2 µl (diluted PCR products) was added to 0.2 µL LIZ 500 size standard and 7.98 Hi-Di Formamide (Applied Biosystems) and separated by capillary electrophoresis using an ABI Prism 3730xl DNA analyzer (Applied Biosystems).

Automated liquid handling systems (Beckman Coulter FX) and bar codes were used to handle the samples during this phase to reduce the occurrence of human errors.

Out of the total sample, 32 accessions, selected amongst those largely known to maximise genetic diversity, were genotyped a second time using a different DNA source, two different polymerases and different cycling profiles and thermal cyclers to verify the reproducibility of profiles and to provide readers with a set of reference profiles.

Analysis of data

Amplicons were analysed and sized with Peak Scanner Software version 1.0 (Applied Biosystems). After a preliminary survey of the data, accessions were reduced to 1005 by eliminating samples that produced large numbers of unreadable profiles, probably due to low DNA quality. The remaining pherograms were all visually inspected, with special attention to potential dropout of long alleles and binning problems.

Data were analysed with the software CERVUS (Marshall et al. 1998; Kalinowski et al. 2007; http://www.fieldgenetics.com) and rare alleles occurring ≤3 times were checked with a second visual inspection of the original pherograms. Genotyping errors were corrected accordingly. Genotypes showing a single peak at a given locus were recorded as homozygous. As a consequence, the reported heterozygosity was underestimated where null alleles occurred at high frequency.

The same software was then used to identify synonymous accessions, and those with exact matches were removed from the data set. The remaining accessions with unique profiles were used to calculate the number of alleles/locus (N_o); the effective number of alleles/locus ($N_e = 1/\sum p_i^2$, where p is the frequency of the ith allele); the observed and expected heterozygosity (Nei 1973, 1987); the power of discrimination (Kloosterman et al. 1993); the polymorphic information content (PIC) (Botstein et al. 1980); the probability of identity ($P_{ID} = \sum p_i^4 + \sum\sum (2p_i p_j)^2$ where p_i and p_j are the frequencies of the ith and the jth allele and $i \neq j$); the fitness to Hardy–Weinberg equilibrium, using the observed and the expected allele frequencies resulting from one generation of random mating and the χ^2 test (Wright 1951); and the estimated frequency of null alleles (null) using the CERVUS iterative algorithm based on the observed and expected frequencies (Summers and Amos 1997).

Parentage analysis was conducted using the CERVUS software after a preliminary simulation that allowed the estimation of critical values of likelihood ratios. Mismatches at a maximum of two loci in each trio were allowed to account for genotyping errors, the occurrence of null alleles, and mutations.

On the same reduced data set, an analysis of the principal components was carried out using the software GenAlEx 6.2 (Peakall and Smouse 2006). A cluster

analysis was also carried out using the unweighted pair group method and arithmetic mean (UPGMA) using the Mega 4.1 software (Tamura et al. 2007).

Then, the program STRUCTURE version 2.2 (Pritchard et al. 2000), a model-based Bayesian clustering method, was used to elucidate the genetic structure of data.

STRUCTURE was run five independent times for each K value (range 1–10, burn-in period of 10,000 generations, 10^5 Markov chain Monte Carlo replications). The adopted admixture model estimates the fraction of ancestry from each cluster for each individual (Pritchard et al. 2000) and the analyses were run with the correlated allele frequencies (Falush et al. 2003). Parameters were set to their default values, as advised in the package documentation (Pritchard and Wen 2004). All individuals were treated as having unknown origin (USEPOPINFO = 0). The choice of the appropriate K value was conducted as recommended in Pritchard and Wen (2004). The K value where ln $P(D)$ was maximum was taken into account.

Finally, we assembled a core collection that should represent the entire genetic diversity explored in the study using the in-house designed CoreFinder software, which can be accessed from the IGA Web site at http://www.appliedgenomics.org. The algorithm for extracting the core collection was obtained heuristically solving the (NP-complete) Set-Covering problem. See the tutorial at the Web site for details.

Results

Procedures and protocols

The number of successfully genotyped accessions ranged from 650 for the locus VChr14b to 1000 for the locus VChr1c out of the 1005 analysed. Successful amplifications in more than 90% of genotypes were achieved for 20/34 SSR loci. Reasons for lack of amplification likely include the occurrence of homozygosity at null alleles (the locus VChr14b was one of these cases), insufficient DNA quality, which is critical for the PCR amplification of several markers, and other technical errors, although automation and bar codes were adopted to minimise the sources of error.

Analysis of identity

The analysis of identity carried out with CERVUS (allowing 0–10 mismatches) found the distribution of incremental hits, which is the number of synonymous accession pairs detected with n mismatches minus the number of synonymous accession pairs detected with $n - 1$ mismatches (Fig. 1).

The distribution of these incremental hits shows a decreasing occurrence between zero and three mismatches,

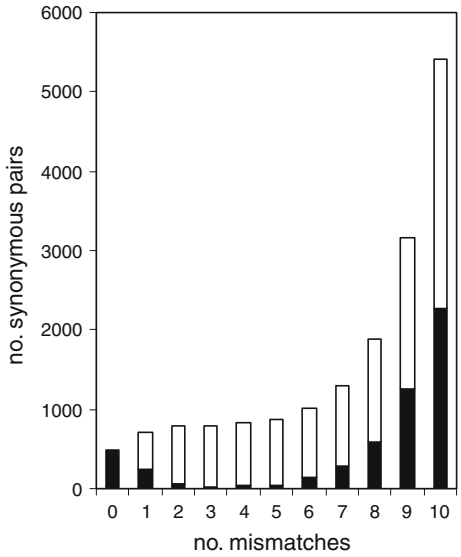

Fig. 1 Absolute (*open bars*) and incremental (*solid bars*) number of synonymous pairs detected at an increasing number of mismatching loci allowed in the accession pair comparison. See "Results" for interpretation of the figure

and an increasing occurrence from four mismatches and higher. The graph continues with a bell-shaped distribution between 5 and 34 mismatches (data not shown). The left part of the histogram should identify the true synonymous accessions, which decrease as the number of allowed mismatches increases. On the other side of the curve, the increment of hits represents pairs of genotypes matching by chance for a population that includes a mixture of full sibs, half sibs, more distant relatives and unrelated genotypes. The hits at the bottom plateau are likely a mixture of the more divergent clonal variants, genotyping errors and full siblings.

Following these observations, we adopted a conservative criterion for the identification of synonyms, accepting two mismatches at most. Using this system, 628 genotypes were identified as unique, whilst 377 genotypes were redundant. The 200 groups of synonymy and near synonymy are reported in the supplementary material S2. If mismatches are not allowed, the number of unique genotypes totals 745.

Many cases of synonymy were expected and several were already reported in the literature, such as 'Greco di tufo' and 'Asprinio' (Costantini et al. 2005). Some others were not, such as the pairs 'Almeria'/'Ciminnita' and 'Marsanne'/'Roussanne', which need to be confirmed by further ampelographic and molecular analyses, granted that the identity of the accessions is correct.

'Leopoldo III', which is molecularly synonymous with 'Alphonse Lavallée', is known to be a tetraploid mutation of the latter, and the two varieties are morphologically different. Many synonymous pairs identified in this study are phenotypically well differentiated from each other for a

number of traits such as skin colour (e.g. the Pinots), leaf morphology (e.g. 'Chasselas dorée' vs. 'Chasselas laciniée'), leaf hairiness (e.g. 'Pinot noir' vs. 'Pinot Meunier'), berry and cluster shape (e.g. 'Catarratto bianco comune' vs. 'Catarratto bianco lucido'), juice aroma (e.g. 'Sauvignon blanc' vs. 'Sauvignon musque') and others.

Several of these traits are important for wine production, and cultivars of these pairs are often worth being individually considered. In the context of the present work, synonymous accessions are entities that can differ from each other in a number of traits. What must be stressed is that these differences originated by mutation, and the synonyms therefore maintain a common genetic origin. In other words, they did not originate through a sexual reproduction.

We also found cases of homonymy, which is defined as an accession classified with the same name but showing a different profile. The accessions of each pair were checked in the field and the one that was incompatible with the identity assigned was considered mislabelled and renamed as 'unknown'. We found 12 of these cases in the repository.

Population statistics of unique accessions

The data set of 745 accessions, which includes near synonyms to avoid missing rare alleles, was used to calculate the population statistics reported in Table 1.

SSR loci were ranked according to their power of discrimination (PD), which is the probability that two randomly sampled accessions in the population of study can be differentiated by their allelic profile at the given locus. Three markers (VChr8b, VChr9b and VChr14b) showed an estimated frequency of null alleles >0.20 and were declassed to the bottom of the list, independently of their PD. However, it is evident that a large positive estimate of null allele frequency simply indicates an excess of homozygotes, which could be expected to some extent in grape, a cleistogamous (selfing) species. Markers showed an average of 8.06 alleles/locus (range 3–21). Most loci showed alleles with extreme frequencies, either low or high, which could explain the low effective number of alleles (N_e = 3.21 on average, range 1.42–6.33). The expected heterozygosity (H_e) was 0.440 over all loci (range 0.204–0.585 for individual loci), whilst the observed (H_o) was lower (0.403) and with a wider range (0.172–0.594). Because genotypes showing a single peak at a given locus were recorded as homozygous, heterozygosity might be underestimated for loci with null alleles occurring at a high frequency. The power of discrimination (PD) was 0.787 over all loci (range 0.459–0.956 for individual loci). The polymorphic information content (PIC) was 0.406 on average (range 0.176–0.572 for individual loci).

A total of 19 loci were in Hardy–Weinberg equilibrium (HW equilibrium), 10 were not, and five were not estimated, mainly because of the high probability of null alleles. The condition of HW equilibrium is an important assumption in the analyses of identity and parentage. Minor deviations from HW equilibrium or deviations at a few loci are not prone to bias likelihood estimates, but deviations at more loci might. Since this is the case in the present study, one should interpret the confidence of identity and parentage assignments with caution. Yet, one should consider that the power of discrimination of the loci in HW equilibrium is high enough to guarantee the robustness of the analysis as a whole.

The probability of identity (P_{ID}) was calculated both for unrelated genotypes and with a more conservative formula for full sibs (Waits et al. 2001). P_{ID} calculated for individual loci ranged from 0.043 for the most discriminating locus VChr3a to 0.539 for the least discriminating VChr12b. The P_{ID} product over all loci was $1.45e^{-27}$ for unrelated genotypes and $9.35e^{-12}$ for full sibs.

A list of the alleles and their relative frequencies is reported for each SSR locus in Table 2. Alleles were called according to the estimated number of repeats contained in the sequence as reported in Cipriani et al. (2008), following the method used in human forensic science.

In Table 3 we reported the profiles of 32 cultivars, which include most alleles (75% of all alleles and 95% of the most frequent alleles, if rare alleles occurring in less than 10 accessions are disregarded). This panel of cultivars can be used as a collection of reference profiles. Several cultivars carry redundant information and have only been included to provide the reader with a larger choice.

Population structure and analysis of principal components

STRUCTURE analysis, which was used to infer K (the number of genetic clusters or gene pools in all accessions), failed to reveal a clear maximum value of ln $P(D)$ over the range of k values, which would allow assignment of the true k value. The 745 cultivars could not be divided into sub-populations, neither when k was left unknown allowing the program to assign it, nor when it was defined 'a priori' on the basis of the geographical origin of accessions (data not shown).

Similar results were obtained with the PCA based on the distance matrix. All methods produced essentially the same pattern. The first three principal axes accounted for 59% of the total variation. Geographic assignments of the accessions were based on either the region of their putative geographic origin or the most prevalent growing area. Seven groups of accessions were established: northeastern Italy (112 accessions), northwestern Italy (83 accessions), central Italy (77 accessions), southern Italy and major islands (91 accessions), northwestern European countries (mainly France, Spain and Germany, 67 accessions), southeastern

Table 1 List of SSR loci ranked according to their power of discrimination

Locus	No. of obs	N_o alleles	N_e alleles	H_o	H_e	PD	PIC	HW	F (null)	P_{ID} unrelated	P_{ID} full sib
VChr3a	655	13	6.33	0.533	0.585	0.956	0.572	***	0.047	0.043	0.340
VChr8a	683	12	5.70	0.418	0.573	0.940	0.556	ND	0.154	0.054	0.351
VChr5a	697	13	5.14	0.554	0.560	0.940	0.544	**	0.007	0.060	0.362
VChr9a	599	10	5.10	0.594	0.559	0.927	0.540	NS	−0.033	0.066	0.364
VCHr5c	568	7	4.21	0.526	0.530	0.906	0.506	NS	0.004	0.091	0.392
VChr18a	719	10	3.64	0.391	0.504	0.882	0.480	***	0.130	0.110	0.415
VChr5b	711	11	3.66	0.524	0.505	0.878	0.475	NS	−0.021	0.117	0.416
VChr19a	719	9	3.16	0.469	0.475	0.866	0.453	NS	0.006	0.131	0.441
VChr15a	703	9	3.31	0.470	0.485	0.866	0.456	NS	0.014	0.133	0.434
VChr13a	664	7	3.18	0.483	0.476	0.861	0.452	NS	−0.008	0.134	0.441
VChr19b	725	5	3.27	0.388	0.482	0.859	0.446	***	0.108	0.145	0.439
VChr7b	639	4	3.31	0.500	0.485	0.851	0.449	NS	−0.018	0.144	0.437
VChr13c	699	4	3.16	0.488	0.475	0.846	0.440	NS	−0.013	0.151	0.446
VChr1a	671	10	2.64	0.446	0.432	0.834	0.415	NS	−0.022	0.167	0.481
VChr13b	710	9	2.53	0.427	0.420	0.816	0.402	NS	−0.008	0.182	0.493
VChr10b	713	5	2.90	0.472	0.455	0.805	0.405	NS	−0.017	0.191	0.470
VChr1b	728	5	2.89	0.468	0.442	0.803	0.403	NS	−0.028	0.188	0.479
VChr18b	739	6	2.88	0.492	0.453	0.795	0.408	*	−0.043	0.186	0.470
VChr12a	732	8	2.62	0.391	0.430	0.778	0.391	***	0.059	0.201	0.491
VChr4a	623	7	2.43	0.416	0.409	0.768	0.369	NS	−0.012	0.227	0.512
VChr16b	707	10	2.11	0.372	0.366	0.749	0.345	NS	−0.008	0.254	0.550
VChr11a	672	7	2.06	0.340	0.358	0.708	0.322	*	0.024	0.287	0.564
VChr14a	535	5	2.32	0.423	0.395	0.704	0.335	NS	−0.040	0.273	0.534
VChr6a	709	4	2.00	0.298	0.347	0.686	0.300	***	0.072	0.318	0.580
VChr2b	645	4	1.80	0.309	0.308	0.648	0.279	NS	−0.001	0.351	0.616
VChr7a	729	3	2.00	0.369	0.347	0.619	0.268	NS	−0.029	0.364	0.591
VChr2a	728	5	1.85	0.383	0.319	0.604	0.267	***	−0.097	0.368	0.612
VChr1c	743	3	1.96	0.378	0.341	0.601	0.263	NS	−0.052	0.372	0.598
VChr17a	651	4	1.76	0.206	0.299	0.597	0.236	***	0.184	0.414	0.638
VChr16a	733	10	1.51	0.172	0.234	0.496	0.223	ND	0.160	0.456	0.696
VChr12b	642	4	1.42	0.212	0.204	0.459	0.176	NS	−0.014	0.539	0.738
SSR with frequency of null alleles >0.20											
VChr8b	680	21	6.06	0.351	0.581	0.941	0.569	ND	0.241	0.044	0.343
VChr9b	513	13	4.38	0.237	0.536	0.867	0.515	ND	0.394	0.082	0.385
VChr14b	495	17	5.88	0.220	0.577	0.899	0.563	ND	0.452	0.048	0.347
Mean[a]	673	8.06	3.21	0.403	0.440	0.787	0.406			1.45E−27	9.35E−12

Loci carrying null alleles at frequencies >0.20 are listed at the bottom

LG linkage group, *No. obs* number of observed accessions ($N_{tot} = 745$), N_o number of alleles, N_e effective number of alleles, H_o observed heterozygosity, H_e expected heterozygosity, *PD* power of discrimination, *PIC* polymorphic information content, *HW* Hardy–Weinberg equilibrium, *F null* estimated frequency of null alleles, *NS* not significant, *significant at the 5% level, **significant at the 1% level, ***significant at the 0.1% level, *ND* not done, P_{ID} probability of identity

[a] In the columns P_{ID}, the mean is substituted with the cumulative P_{ID}, which is the product of the P_{ID} of individual loci

European countries (mainly Balkan countries, 14 accessions) and unknown origin and recent New World cultivars (229 accessions). The 72 table grape cultivars were considered as a separate group and compared with the remaining 673 wine cultivars. Projection of accessions on a two-dimensional plane defined by the first two axes (explaining about 43% of the total variation) showed a slight tendency to separate the table grapes from the wine cultivars (Fig. 2). Projection of accessions on a two-dimensional plane of the six geographically assigned groups showed a slight tendency of northern Italian cultivars to cluster preferentially on the left part of the graph, and the southern Italian

Table 2 List of SSRs in alphabetic order

Locus	Motif	Locus	Alleles												
VChr1a	ATCC	Repeat units	s	0	1	3	6	7	8	9	11	12			
		Length (bp)	175.11	192.61	196.67	205.19	217.70	221.79	226.01	230.22	238.56	242.77			
		SD (bp)	0.16	0.14	0.18	0.19	0.32	0.19	0.23	0.20	0.48	0.26			
		Frequency	0.049	0.001	0.107	0.005	0.009	0.588	0.086	0.077	0.001	0.077			
VChr1b	ATCC	Repeat units	2	3	4	6	7								
		Length (bp)	90.68	94.92	99.16	107.25	111.31								
		SD (bp)	0.20	0.12	0.17	0.16	0.16								
		Frequency	0.262	0.154	0.001	0.066	0.516								
VChr1c	AGCC	Repeat units	3	4	6										
		Length (bp)	90.33	94.69	102.91										
		SD (bp)	0.49	0.11	0.11										
		Frequency	0.404	0.008	0.588										
VChr2a	AGGC	repeat units	3	4	5	7	36								
		Length (bp)	137.06	141.32	146.01	154.54	270.09								
		SD (bp)	0.09	0.28	0.10	0.10	0.63								
		Frequency	0.286	0.004	0.676	0.033	0.001								
VChr2b	AGCT	repeat units	4	5	6	7									
		length (bp)	114.99	119.01	122.89	126.80									
		SD (bp)	0.13	0.11	0.10	0.12									
		Frequency	0.153	0.126	0.719	0.002									
VChr3a	AAT	Repeat units	9	10	11	12	13	14	15	16	17	18	20	24	26
		Length (bp)	174.30	177.59	180.59	183.70	186.70	189.70	192.76	195.79	198.95	202.24	208.17	220.69	226.86
		SD (bp)	0.25	0.17	0.16	0.17	0.12	0.15	0.15	0.14	0.13	0.15	0.14	0.14	0.34
		Frequency	0.005	0.098	0.247	0.105	0.175	0.039	0.011	0.042	0.199	0.004	0.037	0.031	0.008
VChr4a	AAAG	Repeat units	2	4	5	6	7	8	10						
		Length (bp)	173.26	180.81	184.68	188.39	192.15	195.96	203.72						
		SD (bp)	0.13	0.11	0.28	0.09	0.11	0.08	0.08						
		Frequency	0.001	0.282	0.005	0.029	0.074	0.569	0.041						
VChr5a	AGATG	Repeat units	4	6	7	8	11	14	15	16	16 + 3	17	17 + 3	18 + 2	19 + 3
		Length (bp)	182.58	192.68	197.74	202.96	215.59	231.13	235.99	240.40	243.29	245.42	248.68	252.48	258.25
		SD (bp)	0.12	0.10	0.10	0.00	0.14	0.31	0.19	0.29	0.28	0.22	0.26	0.23	0.24
		Frequency	0.160	0.160	0.039	0.001	0.057	0.028	0.001	0.343	0.001	0.019	0.001	0.068	0.123
VChr5b	AAAG	Repeat units	3	4	5	6	6 + 2	7	8	9	10	11	13		
		Length (bp)	178.94	182.88	186.72	190.67	192.59	194.46	198.25	202.22	206.40	210.42	218.42		
		SD (bp)	0.20	0.10	0.24	0.11	0.18	0.08	0.08	0.08	0.05	0.08	0.17		
		Frequency	0.002	0.180	0.015	0.004	0.021	0.399	0.276	0.034	0.004	0.063	0.001		
VChr5c	ACAT	Repeat units	s	2	3	4	6	7	8						
		Length (bp)	83.52	99.94	103.74	107.58	115.48	119.41	123.42						
		SD (bp)	0.17	0.15	0.16	0.14	0.13	0.11	0.12						
		Frequency	0.347	0.049	0.077	0.130	0.018	0.091	0.289						
VChr6a	AATC	Repeat units	3	4	5	6									
		Length (bp)	172.44	176.25	180.11	184.10									
		SD (bp)	0.17	0.20	0.13	0.12									
		Frequency	0.078	0.003	0.649	0.271									
VChr7a	AAAAG	Repeat units	3	5	6										
		Length (bp)	125.58	135.27	140.38										
		SD (bp)	0.08	0.08	0.11										
		Frequency	0.012	0.418	0.570										

Table 2 continued

Locus	Motif	Locus	Alleles												
VChr7b	ACAT	Repeat units	5	7	8	9									
		Length (bp)	171.79	179.94	183.81	187.83									
		SD (bp)	0.42	0.22	0.23	0.26									
		Frequency	0.106	0.425	0.224	0.245									
VChr8a	AAT	Repeat units	11	12	13	14	15	15 + 1	17	17 + 1	18 + 1	19	21	22	
		Length (bp)	171.40	174.67	177.57	180.60	183.65	184.98	189.78	191.13	194.15	195.92	202.50	205.29	
		SD (bp)	0.16	0.17	0.13	0.14	0.13	0.12	0.19	0.11	0.10	0.12	0.07	0.12	
		Frequency	0.064	0.045	0.020	0.220	0.038	0.004	0.009	0.006	0.203	0.157	0.004	0.231	
VChr8b	AAG	Repeat units	8	10	11	12	13	14	15	16	17	18	19	20	21
		Length (bp)	96.24	102.17	105.07	108.04	110.98	113.99	116.94	119.96	122.99	126.10	129.14	132.30	135.38
		SD (bp)	0.15	0.14	0.29	0.13	0.14	0.13	0.10	0.13	0.15	0.13	0.16	0.16	0.10
		Frequency	0.055	0.002	0.306	0.073	0.054	0.004	0.005	0.204	0.113	0.006	0.008	0.046	0.006
		Repeat units	22	23	25 + 1	25 + 2	26 + 1	27 + 1	28 + 1	29 + 1					
		Length (bp)	138.47	141.97	148.74	149.90	151.99	155.02	157.95	160.80					
		SD (bp)	0.07	0.12	0.13	0.15	0.12	0.10	0.00	0.00					
		Frequency	0.037	0.019	0.036	0.008	0.012	0.003	0.001	0.002					
VChr9a	AAG	Repeat units	5	7	8	9	10	11	12	13	14	15			
		Length (bp)	90.03	96.22	99.18	102.19	105.31	108.11	111.10	114.10	117.07	120.14			
		SD (bp)	0.17	0.21	0.15	0.20	0.15	0.11	0.12	0.11	0.10	0.13			
		Frequency	0.032	0.003	0.001	0.214	0.174	0.114	0.040	0.299	0.001	0.124			
VChr9b	AAT	Repeat units	4	5	6	7	8	9	11	12	14	17	18	19	23
		Length (bp)	101.64	104.79	107.55	110.40	113.37	116.34	122.31	125.26	131.35	140.57	144.15	147.51	159.82
		SD (bp)	0.14	0.16	0.14	0.14	0.13	0.12	0.13	0.15	0.13	0.26	0.11	0.00	0.14
		Frequency	0.002	0.007	0.085	0.034	0.271	0.001	0.059	0.031	0.053	0.001	0.362	0.002	0.093
VChr10b	AAC	Repeat units	6	8	9	10	12								
		Length (bp)	120.26	125.97	128.90	131.79	137.68								
		SD (bp)	0.07	0.09	0.09	0.09	0.08								
		Frequency	0.004	0.004	0.429	0.318	0.245								
VChr11a	AAAG	Repeat units	s	4	6	7	7 + 3	8	9						
		Length (bp)	115.46	185.78	194.04	198.16	201.39	202.36	206.62						
		SD (bp)	0.11	0.13	0.11	0.11	0.09	0.13	0.13						
		Frequency	0.004	0.232	0.010	0.653	0.001	0.032	0.068						
VChr12a	AATT	Repeat units	3	4	5	6	7	8	8 + 1	9 + 1					
		Length (bp)	121.13	125.18	129.31	133.49	136.62	140.91	142.08	145.60					
		SD (bp)	0.12	0.10	0.10	0.12	0.11	0.12	0.10	0.16					
		Frequency	0.012	0.001	0.025	0.292	0.053	0.535	0.080	0.003					
VChr12b	AATT	Repeat units	s	4	5	6									
		Length (bp)	122.44	161.24	165.13	169.47									
		SD (bp)	0.18	0.24	0.66	0.11									
		Frequency	0.001	0.173	0.004	0.822									
VChr13a	AAAAG	Repeat units	4	5	5 + 3	6	7	8	10						
		Length (bp)	135.26	140.33	143.63	145.74	150.98	155.84	165.35						
		SD (bp)	0.09	0.08	0.09	0.16	0.06	0.08	0.07						
		Frequency	0.053	0.207	0.114	0.081	0.498	0.023	0.023						
VChr13b	AAAT	Repeat units	4	5	6	6 + 2	7	7 + 1	8	9	10				
		Length (bp)	143.81	148.37	152.65	154.95	156.75	157.80	160.54	164.83	168.99				
		SD (bp)	0.11	0.14	0.13	0.16	0.13	0.15	0.30	0.17	0.39				
		Frequency	0.100	0.121	0.077	0.006	0.601	0.004	0.011	0.030	0.051				

Table 2 continued

Locus	Motif	Locus	Alleles												
VChr13c	AAT	Repeat units	4	6	7	11									
		Length (bp)	114.33	120.24	123.19	135.31									
		SD (bp)	0.13	0.12	0.10	0.10									
		Frequency	0.129	0.163	0.461	0.246									
VChr14a	AATC	Repeat units	s	4	5	19	20								
		Length (bp)	109.61	127.27	131.46	187.73	191.98								
		SD (bp)	0.18	0.41	0.4	0.4	0.5								
		Frequency	0.007	0.078	0.529	0.381	0.005								
VChr14b	ATC	Repeat units	8	9	10	11	12	13	14	15	16	17	20	21	22
		Length (bp)	172.29	175.37	177.65	181.23	183.91	186.71	190.01	192.27	196.03	199.05	208.24	211.34	214.42
		SD (bp)	0.30	0.22	0.51	0.32	0.35	0.20	0.27	0.22	0.18	0.18	0.13	0.13	0.15
		Frequency	0.034	0.003	0.275	0.035	0.226	0.001	0.174	0.078	0.033	0.026	0.023	0.031	0.032
		Repeat units	23	24	25	29	30								
		Length (bp)	217.46	220.55	223.63	235.81	238.92								
		SD (bp)	0.19	0.18	0.04	0.20	0.25								
		Frequency	0.005	0.002	0.016	0.004									
VChr15a	ATCC	Repeat units	2	4	5	6	7	8	9	10	11				
		Length (bp)	127.34	135.47	139.69	144.19	148.67	152.84	156.86	160.68	164.68				
		SD (bp)	0.10	0.10	0.09	0.10	0.11	0.12	0.16	0.11	0.11				
		Frequency	0.021	0.001	0.080	0.136	0.458	0.256	0.001	0.020	0.028				
VChr16a	AAAT	Repeat units	3	5	6	8	13	15	16	18	20	21			
		Length (bp)	99.48	106.77	110.64	118.43	138.40	147.33	151.66	159.40	167.37	171.14			
		SD (bp)	0.15	0.20	0.12	0.12	0.13	0.00	0.00	0.11	0.02	0.00			
		Frequency	0.057	0.807	0.088	0.003	0.001	0.001	0.005	0.005	0.001	0.033			
VChr16b	AATT	Repeat units	2	4	5	5 + 3	6 + 3	7	7 + 3	8	9	10			
		Length (bp)	164.89	172.92	176.94	179.84	183.90	184.91	187.92	188.95	192.97	196.98			
		SD (bp)	0.13	0.16	0.38	0.15	0.12	0.14	0.13	0.12	0.11	0.13			
		Frequency	0.165	0.001	0.022	0.046	0.044	0.018	0.001	0.663	0.038	0.002			
VChr17a	AACC	Repeat units	s	2	5	7									
		Length (bp)	158.69	165.99	176.61	184.41									
		SD (bp)	0.23	0.22	0.14	0.11									
		Frequency	0.002	0.001	0.308	0.689									
VChr18a	AAGG	Repeat units	s	4	5	6	7	8	9	10	11	14			
		Length (bp)	132.94	151.30	155.43	159.44	163.43	167.43	171.43	175.42	179.46	191.52			
		SD (bp)	0.15	0.11	0.12	0.13	0.14	0.15	0.15	0.14	0.26	0.14			
		Frequency	0.001	0.109	0.015	0.232	0.442	0.106	0.031	0.017	0.006	0.041			
VChr18b	AGGC	Repeat units	2	3	4	4 + 1	5	6							
		Length (bp)	136.95	141.50	145.89	147.08	150.34	154.51							
		SD (bp)	0.07	0.21	0.10	0.10	0.10	0.09							
		Frequency	0.001	0.001	0.453	0.177	0.038	0.331							
VChr19a	AAG	Repeat units	5	6	8	9	10	11	12	13	14				
		Length (bp)	123.32	126.40	132.60	135.68	138.80	142.33	145.70	149.05	152.24				
		SD (bp)	0.17	0.15	0.19	0.15	0.16	0.14	0.14	0.14	0.15				
		Frequency	0.020	0.077	0.003	0.035	0.162	0.143	0.511	0.015	0.034				
VChr19b	AGAT	Repeat units	4 + 2	5	6	7	8								
		Length (bp)	156.23	158.33	162.34	166.37	170.37								
		SD (bp)	0.13	0.14	0.16	0.14	0.14								
		Frequency	0.010	0.126	0.431	0.290	0.143								

Alleles (no. of repeats), length (bp) and frequency are reported. Frequencies were calculated after all duplications were removed (*s* small allele, which would indicate the no. of repeats is <0)

Table 3 Profile at 35 SSR microsatellite loci of 32 reference genotypes

Cultivar	VChr1a	VChr1b	VChr1c	VChr2a	VChr2b	VChr3a	VChr4a	VChr5a	VChr5b	VChr5c	VChr6a	VChr7a
Aglianico	7–8	4–6	6–6	5–5	6–6	11–11	7–8	6–14	7–7	2–6	5–6	6–6
Albana	7–7	4–6	4–4	5–7	4–6	11–17	4–8	4–17	8–13	2–2	5–6	6–6
Cabernet franc	7–11	4–6	3–4	5–5	6–6	11–11	4–8	11–18+2	8–11	2–8	5–6	5–6
Cabernet Sauvignon	7–7	4–4	4–4	3–5	6–6	11–17	4–4	11–16	8–11	s–2	5–5	5–5
Chardonnay	7–12	4–7	4–6	5–5	4–6	17–17	4–6	16–18+2	4–7	2–2	3–5	5–6
Chasselas blanc	7–7	4–7	4–6	3–5	5–6	13–17	8–8	6–19+3	7–7	2–6	5–6	6–6
Corinto nero	8–8	4–6	4–4	4–5	4–6	12–13	4–8	7–16	4–7	s–2	5–6	5–6
Dattier de Beyrouth	7–7	4–4	4–4	3–5	5–6	24–24	4–4	4–16	6+2–7	s–4	5–5	5–5
Franconia	1–1	7–7	6–6	3–5	6–6	11–11	4–4	16–18+2	7–8	s–2	3–5	6–6
Garganega	1–7	4–6	4–6	3–5	4–6	11–12	4–8	4–16	7–8	s–2	5–5	5–6
Garnaccia tinta	1–7	6–6	6–6	5–5	6–6	17–20	7–7	16–18+2	4–7	s–s	5–5	5–5
Greco di Tufo	7–8	4–4	4–6	3–5	6–6	11–14	8–8	6–16	4–8	s–2	5–5	5–6
Harslevelu	6–12	3–7	4–6	5–5	4–4	13–17	4–8	6–18+2	4–7	s–2	5–5	5–6
Italia	1–8	4–6	4–4	5–5	5–6	10–11	4–7	7–16	8–8	s–2	3–6	5–5
Malvasia delle Lipari	7–8	4–4	4–6	3–5	6–6	11–13	8–8	4–16	7–8	s–6	5–5	5–6
Malvasia istriana	s–12	4–7	6–6	3–5	4–6	10–10	8–8	16–17	11–11	2–2	5–5	6–6
Mueller Thurgau	7–12	4–7	4–6	5–5	4–6	17–20	8–8	16–18+2	4–7	2–8	3–6	6–6
Muscat blanc à petits grains	s–9	4–7	4–6	3–5	4–5	13–16	4–8	16–16	4–4	2–6	5–6	5–6
Petit Verdot	7–7	4–4	4–4	3–5	4–6	10–17	8–10	14–16	4–11	2–2	5–6	6–6
Primitivo	7–8	6–7	6–6	3–5	6–6	11–13	6–8	16–19+3	7–11	2–8	5–5	5–6
Prosecco tondo	7–7	4–4	4–6	5–5	6–6	10–11	4–8	4–11	7–8	s–2	5–5	5–6
Raboso Piave	8–8	3–4	4–6	5–7	6–6	11–17	7–7	6–16	4–8	6–8	5–6	6–6
Raboso veronese	7–8	4–4	4–6	5–5	6–6	11–11	7–10	6–16	7–8	s–8	5–6	6–6
Refosco dal peduncolo rosso	1–12	4–4	4–6	3–5	5–6	10–17	8–8	14–19+3	7–11	2–8	5–6	5–6
Sangiovese	7–7	4–6	6–6	3–5	5–6	16–17	8–8	18+2–19+3	7–7	s–2	5–5	6–6
Sauvignon	7–7	4–4	4–6	3–5	6–6	10–17	4–8	16–16	7–8	s–2	5–5	5–6
Sultanina	9–9	4–4	4–4	3–5	6–6	15–26	6–8	6–16	7–11	s–7	5–5	3–6
Syrah	7–7	4–7	4–6	3–5	5–5	10–17	4–8	16–19+3	7–8	2–6	6–6	5–6
Traminer rot	7–7	4–6	4–4	3–5	5–6	17–24	4–8	16–16	4–7	2–2	5–5	6–6
Trebbiano toscano	7–7	4–4	4–6	5–5	4–6	12–12	4–8	16–16	7–8	2–3	5–5	5–5
Veltliner rot	1–7	3–4	4–6	3–5	6–6	14–17	4–8	16–19+3	7–8	s–2	5–5	5–6
Verduzzo friulano	7–7	4–7	6–6	3–5	6–6	10–16	4–8	14–19+3	7–11	2–8	5–5	6–6

Cultivar	VChr7b	VChr8a	VChr8b	VChr9a	VChr9b	VChr10b	Vchr11a	Vchr12a	Vchr12b	Vchr13a	Vchr13b	Vchr13c
Aglianico	5–7	22–22	12–26+1	5–12	5–6	9–12	7–7	8–8	4–4	7–7	5–7	11–11
Albana	7–9	22–22	13–13	5–14	8–14	9–12	7–7	8–8	4–6	5–7	7–10	7–7
Cabernet franc	7–9	18+1–22	10–10	5–15	8–18	9–9	4–7	6–6	4–6	5–7	5–6	4–7
Cabernet Sauvignon	8–9	18+1–22	10–12	5–8	8–14	9–10	4–9	6–8	4–6	5–5	5–6	4–7
Chardonnay	9–9	18+1–22	21–21	5–8	8–14	9–12	7–7	6–8	6–6	7–7	7–7	6–7
Chasselas blanc	5–9	18+1–19	12–12	8–8	6–14	10–12	7–7	6–6	6–6	7–7	6–7	4–7
Corinto nero	7–7	12–17+1	10–10	9–12	6–8	9–9	7–7	6–8	6–6	4–5+3	4–7	7–7
Dattier de Beyrouth	7–8	18+1–18+1	10–18	5–12	8–8	10–12	4–7	6–8	6–6	5+3–8	4–5	7–11
Franconia	9–9	14–22	17–17	5–14	7–8	9–9	7–9	7–8	6–6	7–7	7–7	7–7
Garganega	7–7	18+1–22	12–13	12–14	7–8	10–12	7–7	6–8	4–6	5–7	7–7	6–7
Garnaccia tinta	7–7	18+1–18+1	14–14	5–5	7–8	9–12	7–7	6–8	4–6	5–10	7–7	4–11
Greco di Tufo	5–9	14–22	14–17	9–12	6–14	9–12	7–9	8–8	4–6	7–8	4–6	11–11
Harslevelu	7–9	18+1–18+1	13–18	5–12	8–8	10–10	7–7	8–8	6–6	7–10	5–7	7–7
Italia	7–7	22–22	10–12	5–5	6–11	9–10	7–7	6–8+1	6–6	6–7	6–7	6–11
Malvasia delle Lipari	7–8	18+1–18+1	24–24	13–14	6–14	9–12	7–7	6–8	6–6	5+3–10	4–9	4–6
Malvasia istriana	5–9	22–22	10–21	12–13	8–14	9–9	4–7	6–8	6–6	6–7	7–7	7–7

Table 3 continued

Cultivar	VChr7b	VChr8a	VChr8b	VChr9a	VChr9b	VChr10b	Vchr11a	Vchr12a	Vchr12b	Vchr13a	Vchr13b	Vchr13c
Mueller Thurgau	9–9	11–17	21–21	8–13	18–23	12–12	7–9	6–7	6–6	6–7	7–7	4–7
Muscat blanc à petits grains	7–7	19–19	12–23	5–13	7–7	9–10	7–7	8–8	6–6	7–7	6–6	11–11
Petit Verdot	5–7	12–19	20–20	5–8	6–6	9–12	4–7	6–6	5–6	4–6	7–9	6–11
Primitivo	7–8	12–12	11–18	5–14	8–8	9–10	6–7	6–7	6–6	6–8	7–7	7–11
Prosecco tondo	5–8	14–14	17–17	13–14	7–14	9–10	7–9	8–8	6–6	5–5+3	5–7	6–7
Raboso Piave	5–7	11–14	17–25+1	12–14	14–18	12–12	4–4	6–8	6–6	6–7	7–7	4–7
Raboso veronese	5–5	14–22	25+1–25+1	5–14	6–18	9–12	4–7	6–8	6–6	7–7	7–7	4–11
Refosco dal peduncolo rosso	7–9	14–19	12–14	8–14	18–18	9–12	7–7	6–8+1	6–6	4–7	4–10	11–11
Sangiovese	9–9	11–14	17–17	5–13	12–12	10–12	7–7	8–8	6–6	7–7	7–7	6–11
Sauvignon	7–8	18+1–19	12–14	8–8	14–18	9–10	7–9	8–8	6–6	5–7	6–7	6–7
Sultanina	7–8	18+1–19	10–12	12–12	11–11	9–10	7–7	6–6	4–6	5–10	7–7	4–7
Syrah	8–9	11–18+1	12–12	13–14	18–18	10–12	4–7	7–8+1	6–6	7–7	7–7	7–11
Traminer rot	7–9	12–19	12–27+1	5–8	18–18	9–10	9–9	6–8	6–6	4–7	6–7	4–6
Trebbiano toscano	7–9	18+1–22	12–17	8–14	8–14	10–12	7–7	8–8	4–6	5+3–7	6–7	6–6
Veltliner rot	7–8	12–14	12–18	5–15	8–14	9–10	7–9	6–8	6–6	7–7	5–7	6–6
Verduzzo friulano	9–9	14–19	14–18	5–5	5–8	9–10	7–7	3–6	6–6	5–5+3	4–10	4–7

Cultivar	Vchr14a	Vchr14b	Vchr15a	Vchr16a	Vchr16b	Vchr17a	Vchr18a	Vchr18b	Vchr19a	Vchr19b
Aglianico	5–19	13–13	5–8	6–6	6+3–7	5–5	5–6	4–4	11–12	7–7
Albana	4–5	11–11	7–8	6–6	5+3–8	5–5	7–7	4–6	10–11	6–6
Cabernet franc	5–19	15–23	7–8	6–6	2–8	7–7	6–9	4–4	11–12	6–6
Cabernet Sauvignon	5–19	11–15	7–8	6–6	2–8	7–7	7–9	4–4	11–12	6–7
Chardonnay	19–19	11–11	2–7	18–20	2–7+3	7–7	7–7	4–4	5–11	7–7
Chasselas blanc	5–5	11–16	7–8	3–6	2–8	5–7	7–9	4+1–4+1	6–10	7–8
Corinto nero	5–5	11–11	7–8	5–6	7+3–8	5–5	6–6	6–6	12–14	6–8
Dattier de Beyrouth	5–5	17–17	5–7	6–6	8–8	5–7	6–6	4–6	11–12	4+2–6
Franconia	19–19	15–17	2–7	6–6	2–2	7–7	4–7	4–4	5–9	6–6
Garganega	4–5	13–15	6–8	6–8	5+3–8	5–5	7–7	4–6	11–12	7–7
Garnaccia tinta	5–19	11–11	6–7	6–8	8–8	7–7	7–7	4+1–5	12–14	6–6
Greco di Tufo	5–5	13–13	6–11	6–6	8–8	5–5	6–7	4–6	11–12	5–7
Harslevelu	5–19	17–17	7–10	6–8	5+3–8	5–5	7–14	4–6	12–12	5–5
Italia	4–19	11–11	8–8	6–8	8–8	5–5	4–6	5–6	12–14	6–8
Malvasia delle Lipari	5–19	15–15	7–7	6–6	7–8	5–7	4–7	4–6	6–10	5–7
Malvasia istriana	5–19	11–11	6–8	6–6	2–9	7–7	4–8	4–4	11–12	5–6
Mueller Thurgau	19–19	11–11	7–8	3–8	8–8	7–7	6–7	4–6	9–11	7–7
Muscat blanc à petits grains	19–19	11–11	7–8	6–6	7+3–9	5–7	8–8	4–6	10–10	6–6
Petit Verdot	5–19	10–11	7–8	3–6	2–2	7–7	7–8	4–6	10–12	7–8
Primitivo	19–19	9–11	7–7	6–6	8–8	7–7	7–14	6–6	9–10	6–6
Prosecco tondo	5–19	11–11	7–7	6–6	8–8	7–7	7–14	4+1–6	11–12	6–8
Raboso Piave	5–19	13–13	7–8	6–8	2–8	7–7	7–7	4–6	10–12	7–7
Raboso veronese	5–19	9–9	7–8	6–8	8–8	7–7	4–7	4–6	10–11	7–7
Refosco dal peduncolo rosso	4–19	11–11	10–11	6–6	8–8	7–7	7–7	4–6	12–12	6–6
Sangiovese	5–5	11–11	6–8	6–6	8–8	7–7	7–8	4–4+1	6–12	7–7
Sauvignon	5–19	11–12	7–7	6–6	8–8	7–7	7–7	4–4+1	6–12	7–7
Sultanina	5–5	22–22	6–6	8–8	5+3–8	7–7	6–11	5–6	10–12	5–6
Syrah	5–5	11–11	7–7	6–6	2–8	7–7	7–7	4–6	6–12	7–7
Traminer rot	5–19	11–11	7–7	6–6	8–8	7–7	4–7	4–4+1	6–9	7–7
Trebbiano toscano	5–5	9–13	6–7	6–6	5+3–8	5–5	7–7	4–4+1	12–12	6–7
Veltliner rot	19–19	13–15	7–7	6–6	7+3–8	7–7	4–7	4–6	13–13	5–5
Verduzzo friulano	5–5	11–11	7–7	6–8	5+2–5+2	7–7	4–7	4–4+1	10–11	6–7

Data represent the number of repeats in the microsatellite. '+n' means additional 'n' bases to the repeats; alleles of the same locus are separated by '–'; profiles with a single peak are conventionally considered homozygous

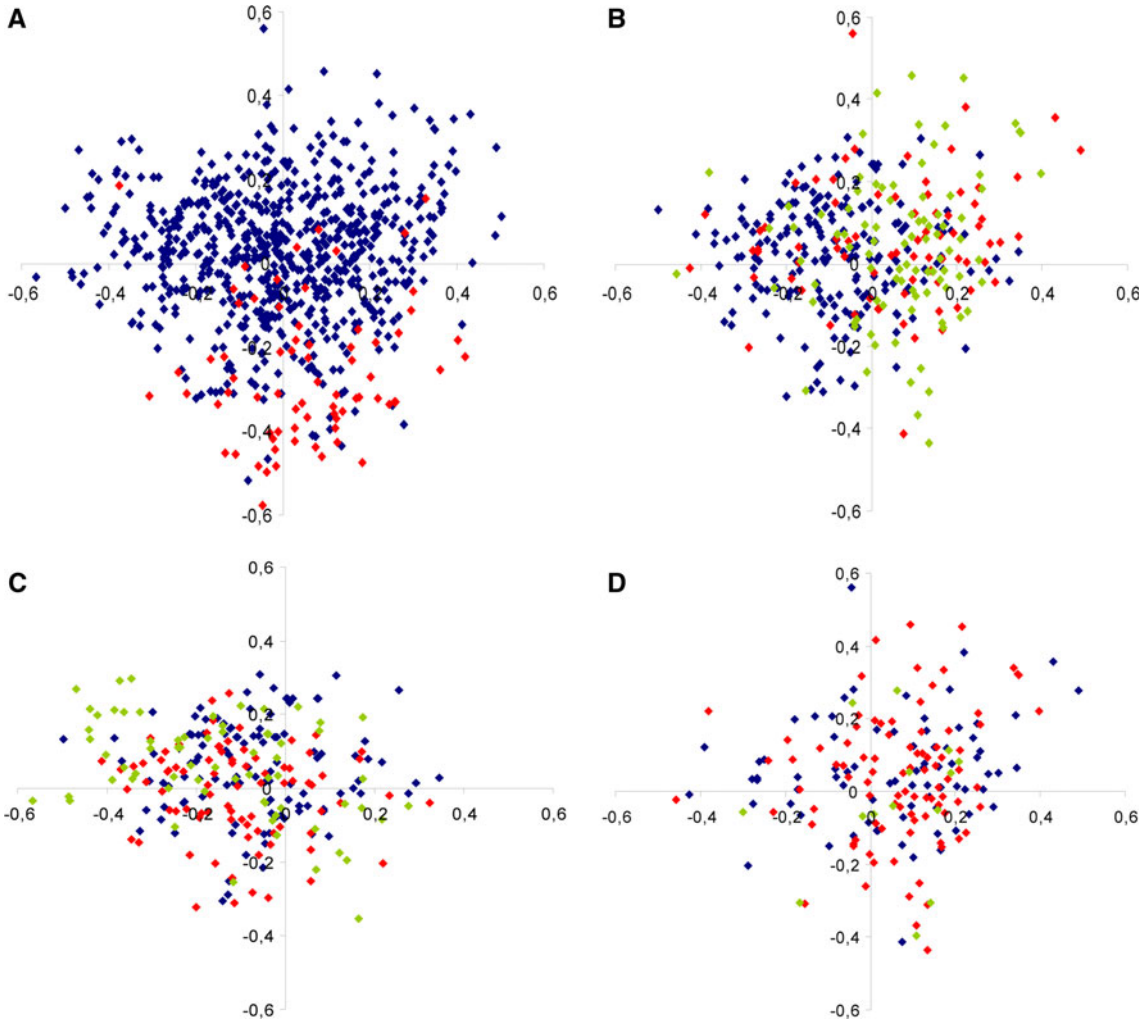

Fig. 2 Principal component analysis of 745 unique accessions of *Vitis vinifera*. The *four graphs* represent the two-dimensional projections of accessions along the first two principal axes accounting for about 43% of the total variation. **a** Reports the distribution of the wine cultivars (*blue diamond*) and table cultivars (*red diamond*), respectively. **b** Reports the distribution of the northern (*blue diamond*), central (*red diamond*) and southern Italian accessions (*green diamond*), respectively. **c** The distribution of the northeastern (*blue diamond*) and northwestern Italian cultivars (*red diamond*), and the northwestern international cultivars (*green diamond*), respectively. **d** Distribution of the central (*blue diamond*) and southern Italian cultivars (*red diamond*), and southeastern international cultivars (*green diamond*), respectively. The accessions for which geographic data were not known were not plotted

cultivars on the right (Fig. 2). The northwestern and northeastern Italian cultivars did not show any particular cluster and could not be separated. The northwestern international cultivars showed a similar pattern, clustering on the upper left part of the two-dimensional plane. The southern Italian cultivars showed the same pattern as the southeastern European cultivars, clustering on the right part of the graph, even though only few accessions were available from that area. The cultivars of unknown origin clustered randomly on the graph (data not shown).

The UPGMA cluster analysis of the distance matrix between genotypes confirmed the picture of overall relationships (Supplementary material S3). There was no evidence of any hierarchical levels at which clusters could be recognised on the pherogram. Accessions belonging to the different geographic areas appeared to be intermixed, and no sub-population could be recognised.

Pedigree analysis

The cases of parent–offspring relationships were explored with the CERVUS software, which was adopted because (a) it allows missing values and mismatches (Kalinowski et al. 2007) and, although being written for the genetic analysis of animal populations, it allows for selfing. The analysis was carried out on the reduced database of 745 accessions, after the perfect synonyms were removed. Near synonymous accessions were maintained in the data set to improve the chance of finding the true parents. We imposed the search for both parents and included the check for

selfing. After a first run, the analysis of data revealed that most mismatches occurred at locus VChr9b because of its high frequency of null alleles. This locus was therefore removed and the analysis was repeated.

Many of the resulting trios consisted of parent–offspring relationships between synonymous and/or near synonymous accessions, as expected. After these false trios were removed, a total of 211 putative trios (offspring and both parents or one parent in cases of selfing) remained. Of these, 94 were indicated by the program with high confidence (95%), 19 with relaxed confidence (80%) and the remaining 98 without an assigned confidence level, usually because the program was not able to select one parent of the pair amongst a number of candidates with equal probability. Pedigrees with incompatible profiles at one or two SSR loci were manually checked. In most cases, the incompatible profiles occurred at loci with a high frequency of null alleles, and thus could be solved with the hypothesis of the occurrence of a null allele in either parent or its offspring.

The most interesting parentages were confirmed with supplementary data already available at the CRA-VIT molecular database at 7–30 further SSR loci, most of which came from the list of di-nucleotide repeats reported in literature (This et al. 2004; Di Vecchi Staraz et al. 2007). Relevant parentages are reported in Supplementary material S4 with reference literature where available.

Many known parentages were confirmed and do not need any further comment, apart from the controversy on the direction of the parent–offspring relationship in the 'Ciliegiolo/Sangiovese' pair. Our data confirm that 'Ciliegiolo' is the offspring of 'Sangiovese' × 'Moscato violetto' (also known as 'Muscat rouge de Madere') in agreement with Di Vecchi Staraz et al. (2007) and in disagreement with Vouillamoz et al. (2007).

Nine parentages were completed or revised. For instance, 'Lagrein', which is a variety grown in South Tyrol (Italy), was previously reported as a likely offspring of 'Teroldego', with the second parent unknown (Vouillamoz and Grando 2006). We were able to confirm 'Teroldego' as one parent and to identify 'Schiava gentile', a variety grown in the same area, as the second parent. The Manzoni crosses 2–14 and 2–15 that were reported by the breeder as 'Cabernet Sauvignon × Prosecco' actually have 'Cabernet franc' instead of 'Cabernet Sauvignon' as a parent. 'Covè', reported as a cross of 'Harslevelu × Malvasia bianca lunga' resulted instead from a selfing of 'Harslevelu'.

Forty parentages have never been reported before. They include several well-known varieties and many less known varieties, as well as varieties grown in very confined geographic areas. A new relevant pedigree is one involving 'Chenin blanc', an old French variety that has been widely distributed in the New World. 'Chenin blanc' is likely the result of the cross 'Sauvignon x Rot Traminer'.

'Muscat of Alexandria', one of the founders of the Muscat family (Crespan and Milani 2001), surprisingly resulted from the cross 'Muscat blanc à petits grains × Axina de tres bias'. This new pedigree is supported by the profile at 61/63 SSR loci and, if it is true, 'Muscat blanc à petits grains' remains the only known founder of the Muscat family carrying the typical aroma. The second parent, 'Axina de tres bias' is an ancient black berry variety still grown on the island of Sardinia, which was never suspected to be involved in Muscat pedigrees. It is also known as 'Trifera', 'Uva di tre volte' and 'Tre volte l'anno', names which refer to its ability to flower and crop three times a year under favourable climatic conditions. This strange behaviour was previously described by Gallesio and Di Rovasenda (cited in Galet 2000). In spite of its lack of any agronomic interest, it became widespread in the past and is still grown in Sardinia and Sicily. According to Logothetis, it is also largely known in the Greek islands and very recently it has also been found on the island of Malta (Crespan et al. 2010 AJEV in publication). In addition, other relationships within the complex Muscat family have been discovered. A parentage three is reported in Fig. 3 for the convenience of the reader. 'Moscatello selvatico' is a cross of 'Muscat blanc à petits grains' × 'Bombino bianco', whilst 'Moscato violetto' is confirmed as a cross of 'Muscat blanc à petits grains' × 'Duraguzza', the second parent being synonymous with 'Mammolo' reported in Di Vecchi Staraz et al. (2007). However, the correct identity of 'Mammolo' is far from being established, given that five genetically distinct grapevine varieties share the same name. Amongst these homonyms, 'Duraguzza'/'Mammolo' encompasses in turn several synonyms found in different Italian regions (Emilia Romagna, Sardinia, Puglia). It seems therefore to be a key genotype of the Italian germplasm since it appears in many different known pedigrees, amongst which is cited 'Moscato violetto'.

We were able to identify the first offspring of the 'Trebbiano toscano' named 'Garofana', a minor variety of the Tuscany germplasm. Other pedigrees involving 'Trebbiano toscano' as a parent still need to be confirmed. If so, they should be the expected result of the large diffusion of this grapevine variety throughout Italy and France.

A further noteworthy discovery involves the black Malvasias grown in Italy. 'Malvasia nera' of Brindisi/Lecce, which is widespread in southern Italy, resulted from a cross of 'Negroamaro' × 'Malvasia Bianca lunga', whilst 'Malvasia nera' of Bolzano (code no. 128 of the National Catalogue) is a cross of 'Perera' × 'Schiava gentile', two varieties typical of the northern part of Italy, which have no relationship either with the other Malvasias or with Greek or middle eastern varieties.

The varieties 'Gaglioppo' grown in Sicily and 'Nerello mascalese' grown in Calabria, previously reported as belonging to the 'Sangiovese' family (Di Vecchi Staraz

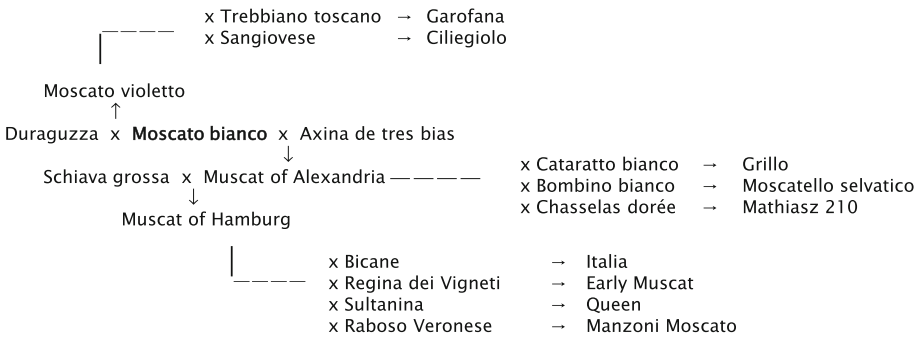

Fig. 3 The new pedigree reconstruction of the Muscat family based on first degree parentages. All parentages are from molecular profiles analysed in the present paper, except for the parentage of 'Muscat of Hamburg', which is from Crespan (2003). The 'Moscato bianco' (syn. 'Muscat blanc à petits grains'), which seems the most ancient Muscat amongst those appearing in the reconstruction, is in *bold*

et al. 2007), were confirmed as offspring of 'Sangiovese', the second parent being 'Mantonico bianco'.

Several known parentages reported in literature (Sefc et al. 1998; Crespan and Milani 2001; Di Vecchi Staraz et al. 2007; Vouillamoz et al. 2007; Bautista et al. 2008; Ibáñez et al. 2009) were not identified, mainly because either parent was not included in the database.

Considering the explorative value of this analysis and the limited number of loci analysed, we neither checked for single parents nor explored more distant relationships.

Sorting out a core collection

A core collection, which is a set of a minimum number of accessions that represents the whole genetic variability of the entire collection, has been extracted without any stratification following the so-called M-method developed by Schoen and Brown (1993) and described by Gouesnard et al. (2001). The 274 alleles identified at the 34 microsatellites used in this study were fully represented by as few as 30 accessions (best of six runs). To demonstrate the power of the algorithm developed in the COREFINDER tool, we report in Fig. 4 the percentage of the total number of alleles as a function of the increasing number of accessions extracted by the program.

The power of the approach based on SSR profiles in sampling a germplasm collection was already demonstrated in grape by Le Cunff et al. (2008), who sampled 271 alleles showing in some cases a frequency as low as 0.05% in their original collection with as few as 48 accessions. The same authors also demonstrated that sampling based on the profile of 20 SSR markers provided a good representation of the genetic diversity found in three gene sequences.

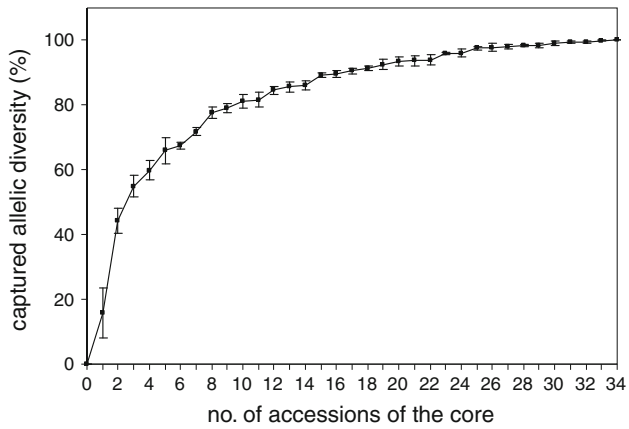

Fig. 4 Genetic diversity (expressed as percentage of the total number of alleles) as a function of the number of accessions included in the core collection. The accessions were re-extracted at each step and they could be different from those extracted in the previous one. Moreover, different iterations of the program could sort out slightly different sets of accessions. *Vertical bars* represent the standard deviation of six different runs

Discussion

Genotyping large collections is challenging because genotyping errors cannot be mitigated by systematically repeating the analysis or using other expedients routinely adopted when the number of samples is more limited (Pemberton et al. 1995; Matsuoka et al. 2002). In spite of this, we gathered a reliable picture of the genetic diversity of the Italian grapevine collection, which includes a good representation of the internationally grown French varieties, the nearly complete National Catalogue and a large sample of local germplasm.

As expected, we found a large occurrence of synonymy and near synonymy. This is likely the result of diffuse exchanges of plant material amongst populations by individuals, as well as the traditional practice of vegetative propagation of grapevine varieties, which led to the selection of variants that arose from natural mutations (Crespan 2004).

If we accept that accessions showing two differences in the profile at 34 SSRs could still be considered synonymous, then up to a 6% mismatch rate could be still accepted for true synonymy, especially for genotypes vegetatively propagated over a long period of time. If genotyping errors

could be excluded from the analysis, the percentage of mismatches should be slightly reduced to about 4%, according to the error estimate reported in human high-throughput genotyping (Ewen et al. 2000).

Siblings could be very frequent in grape, as we noticed from the number of accessions differing from each other at very few loci. Many of these could have originated from selfing, considering that domesticated grape is cleistogamous. This could explain the excess of homozygotes found at several loci, which, in combination with the frequency of null alleles, could be a source of the deviation of those loci from Hardy–Weinberg equilibrium.

Both identity and pedigree analyses need to be confirmed by observations of ampelographic traits, since we cannot exclude cases of further misnaming and mislabelling in the collection. Germplasm curators will perform this task, but we wish to stress here that parentages and kinships not known from historical records can only be discovered through massive genotyping of large collections (Bowers et al. 1999). This is particularly true for wine grape varieties that have been grown for centuries with an unknown origin.

Large databases can also provide information on more relaxed relationships and kin groups (Konovalov et al. 2004, Vouillamoz and Grando 2006, Di Vecchi Staraz et al. 2007), as well as on levels of coancestry and inbreeding (Reynolds et al. 1983). Although they appear relevant in grape, we did not explore them because of the high rate of false positive/false rejection that would likely occur when analysing a small number of loci (Di Vecchi Staraz et al. 2007). Some 60 SSR loci, the number depending on their level of polymorphism and power of discrimination, could be required to discriminate amongst alternative parents that are relatives of each other and to check parent–offspring relationships against kinships of different degrees as suggested by Vouillamoz and Grando (2006).

In spite of the alleged independence of the grapevine gene pools of different European grapevine-growing regions that was initially reported (Sefc et al. 2000), new evidence shows complex relationships of second and third degree between cultivars considered typical of the traditional European growing areas, challenging the independent origin of those cultivars (Vouillamoz and Grando 2006; Di Vecchi Staraz et al. 2007). This is likely the reason for the weak structure found in the germplasm collection analysed in this work.

This study has confirmed the usefulness of the new set of markers for grape genotyping (Cipriani et al. 2008). They offer a good level of polymorphism, are easy to use and score, and alleviate the problems with binning and allele calling produced by the di-nucleotide repeat markers of the past (This et al. 2004). The new method of allele calling, based on the number of repeats in the sequence, aids in the reduction of allele miscalling and definitively aligns grape fingerprinting methods to the protocols used in humans.

The calculation of allele frequencies over a large number of individuals increases the reliability of calculating probability ratios in the identity analysis. The overall probabilities of identity, which are the probabilities that two unrelated or sibling individuals randomly drawn from the population have identical profiles at all loci, are $1.45e^{-27}$ and $9.35e^{-12}$, respectively. The value is extremely low in both cases, especially considering that the total number of grapevine accessions in the world repositories does not likely exceed 10,000–14,000 genotypes, according to different authors (Alleweldt et al. 1990; This et al. 2006), including synonymy, homonymy and misnaming.

The departure from HW equilibrium at several loci may suggest removing them from the set, thus preserving the formal value of the analysis. The remaining loci should be sufficient to guarantee a negligible probability of error, at least for the identification of synonymy. The 13 CODIS loci adopted in human molecular fingerprinting generate profile frequencies on the order of one in tens of billions, or more in cases of less frequent genotypes and alleles. Such ratios are considered sufficiently discriminating in the human population, which totals 6 billion people. We are confident to have reached a comparably low probability of error, at least for unrelated genotypes.

The compilation of a core collection representing the entire allelic richness demonstrated that very few accessions are needed. Genetic diversity was low in our collection, probably because it was restricted to cultivated genotypes of *V. vinifera* subsp. *sativa* of the national collection. Nevertheless, the high level of heterozygosity recorded in most individuals and the expected presence of large kin groups (Crespan et al. 2008a, b) are likely the two other factors underlying the low number of accessions required to extract the core collection.

Acknowledgments This research was financially supported by the Ministry of Agriculture and Forestry, project VIGNA, and by the Istituto di Genomica Applicata (IGA), Parco Scientifico e Tecnologico di Udine. The authors thank Alberto Stefan at IGA, who wrote several scripts for the management of input/output data, and Courtney Coleman at the University of Missouri State University for the critical review of the manuscript. Authors thank also Mirella Giust for field collection of samples, and Massimo Gardiman and Roberto Carraro for the visual inspection of genotypes of dubious identity.

References

Akkak A, Boccacci P, Botta R (2007) 'Cardinal' grape parentage: a case of a breeding mistake. Genome 50:325–328

Alleweldt G, Spiegel-Roy P, Reisch (1990) Grapes (Vitis). In: Moore JN, Ballington JR Jr (eds) Genetic resources of temperate fruit and nut crops, vol I. ISHS, Leuven, pp 291–327

Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS (2007) Automated binning of microsatellite alleles: problems and solutions. Mol Ecol Notes 7:10–14

Arroyo-García R, Ruiz-García L, Bolling L, Ocete R, López MA, Arnold C, Ergul A, Söylemezoğlu G, Uzun HI, Cabello F, Ibáñez J, Aradhya MK, Atanassov A, Atanassov I, Balint S, Cenis JL, Costantini L, Gorislavets S, Grando MS, Klein BY, McGovern PE, Merdinoglu D, Pejic I, Pelsy F, Primikirios N, Risovannaya V, Roubelakis-Angelakis KA, Snoussi H, Sotiri P, Tamhankar S, This P, Troshin L, Malpica JM, Lefort F, Martinez-Zapater JM (2006) Multiple origins of cultivated grapevine (*Vitis vinifera* L, ssp *sativa*) based on chloroplast DNA polymorphisms. Mol Ecol 15(12):3707–3714

Bautista J, Dangl GS, Yang J, Reisch B, Stover E (2008) Use of genetic markers to assess pedigrees of grape cultivars and breeding program selections. Am J Enol Vitic 59(3):248–254

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

Boursiquot JM, Lacombe T, Bowers J, Meredith C (2004) Le Gouais, un cépage clé du patrimoine viticole européen [Gouais, a key vine of the European wine heritage]. Bull l'OIV 77(875–876):5–19

Bowers JE, Meredith CP (1997) The parentage of classic wine grape: Cabernet Sauvignon. Nat Genet 16:84–87

Bowers J, Boursiquot JM, This P, Chu K, Johansson M, Meredith C (1999) Historical genetics: the parentage of Chardonnay, Gamay and other wine grapes of Northeastern France. Science 285:1562–1565

Butler JM (2006) Genetics and genomics of core short tandem repeats loci used in human identity testing. J Forensic Sci 51(2):253–265

Calò A, Scienza A, Costacurta A (2001) Vitigni d'Italia. Calderini Edagricole, Bologna, Italy

Cancellier S, Roncador I (1997) Gli incroci Manzoni. La Tipografica 2000. Treviso, Italy

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

Cipriani G, Di Gaspero G, Canaguier A, Jusseaumes J, Tassin J, Lemainque A, Adam-Blondon AF, Testolin R (2009) Molecular linkage maps: Strategies, resources and achievements. In: Zapater MM, Adam-Blondon AF (eds) Grapes. Series on "Genomics of fruits and vegetables crops", Series Editor: Chittaranjan Kole. Sciences Publishers, Enfield (in press)

Costantini L, Monaco A, Vouillamoz J, Forlani M, Grando MS (2005) Genetic relationships among local *Vitis vinifera* cultivars from Campania (Italy). Vitis 44(1):25–34

Crespan M (2003) The parentage of Muscat of Hamburg. Vitis 42:193–197

Crespan M (2004) Evidence on the evolution of polymorphism of microsatellite markers in varieties of *Vitis vinifera* L. Theor Appl Genet 108:231–237

Crespan M, Milani N (2001) The Muscats: a molecular analysis of synonyms, homonyms and genetic relationships within a large family of grapevine cultivars. Vitis 40(1):23–30

Crespan M, Crespan G, Giannetto S, Meneghetti S, Costacurta A (2007) Vitouska' is the progeny of 'Prosecco tondo' and 'Malvasia bianca lunga. Vitis 46(4):192–194

Crespan M, Calò A, Giannetto S, Sparacio A, Storchi P, Costacurta A (2008a) 'Sangiovese' and 'Garganega' are two key varieties of the Italian grapevine assortment evolution. Vitis 47(2):97–104

Crespan M, Coletta A, Crupi P, Giannetto S, Antonacci D (2008b) Malvasia nera di Brindisi/Lecce' grapevine cultivar (*Vitis vinifera* L.) originated from 'Negroamaro' and 'Malvasia bianca lunga'. Vitis 47(4):205–212

Di Vecchi Staraz M, Bandinelli R, Borselli M, This P, Boursiquot JM, Laucou V, Lacombe T (2007) Genetic structuring and parentage analysis for evolutionary studies in grapevine: kin group and origin of the cultivar Sangiovese revealed. J Amer Soc Hort Sci 132(4):514–524

Ewen KR, Bahlo M, Treloar SA, Levinso DF, Mowry B, Barlow JW, Foote SJ (2000) Identification and analysis of error types in high-throughput genotyping. Am J Hum Genet 67:727–736

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Galet P (2000) Dictionnaire encyclopedique des cépages. Hachette Pratique, Paris, France

Gouesnard B, Bataillon TM, Decoux G, Rozale C, Schoen DJ, David JL (2001) MSTRAT: an algorithm for building germplasm core collections by maximizing allelic or phenotypic richness. J Heredity 92:93–94

Ibáñez J, Vargas AM, Palancar M, Borrego J, de Andrés MT (2009) Genetic relationships among table-grape varieties. Am J Enol Vitic 60(1):35–42

Idury RM, Cardon LR (1997) A simple method for automated allele binning in microsatellite markers. Genome Res 7:1104–1109

Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyere C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pé E, Valle G, Morgante M, Caboche M, Adam-Blondon A-F, Weissenbach J, Quétier F, Wincker P (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–468

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106

Kloosterman AD, Budowle B, Daselaar P (1993) PCR-amplification and detection of the human DIS80 VNTR locus. Amplification conditions, population genetics and application in forensic analysis. Int J Leg Med 105:257–264

Konovalov DA, Manning C, Henshaw MT (2004) KINGROUP: a computer program for pedigree relationship reconstruction and kin group assignments using genetic markers. Mol Ecol Notes 4:779–782

Lacombe T, Boursiquot JM, Dechesne F, Varès D, This P (2007) Relationships and genetic diversity within the accessions related to Malvasia Held in the Domaine de Vassal Grape Germplasm Repository. Am J Enol Vitic 58(1):124–131

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

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 7:639–655

Matsuoka Y, Vigouroux Y, Goodman MM, Jesus Sanchez G, Buckler E, Doebley J (2002) A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA 99:6080–6084

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Olmo HP (1995) Grapes. In: Smartt J, Simmonds NW (eds) Evolution of crop plants, 2nd edn. Longman, London, pp 485–490

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Pemberton JM, Slate J, Bancroft DR, Barrett JA (1995) Non-amplifying alleles at microsatellite loci: a caution for parentage and population studies. Mol Ecol 4:249–252

Pritchard JK, Wen W (2004) Documentation for structure software, Version 2. Department of Human Genetics, University of Chicago, Chicago, USA

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure from multi-locus genotype data. Genetics 155:945–959

Reynolds J, Wier BS, Cockerham CC (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics 105:767–779

Roncador I, Malossini U, Grando MS, Mattivi F, Nicolini G, Versini G (2002) Caratteristiche viti-enologiche dei nuovi vini Goldtraminer, Sennen e Gosen. Terra Trentina 48:28–36

Schlötterer C (2004) The evolution of molecular markers—just a matter of fashion? Nat Rev Genet 5:63–69

Schoen DJ, Brown AHD (1993) Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. Proc Natl Acad Sci USA 90:10623–10627

Sefc KM, Steinkellner H, Glőssl J, Kampfer S, Regner F (1998) Reconstruction of a grapevine pedigree by microsatellite analysis. Theor Appl Genet 97:227–231

Sefc KM, Lopes MS, Lefort F, Botta R, Boubelakis-Angelakis KA, Ibañez J, Pejić I, Wagner HW, Glőssl J, Steinkellner H (2000) Microsatellite variability in grapevine cultivars from different European regions and evaluation of assignment testing to assess the geographic origin of cultivars. Theor Appl Genet 100:498–505

Summers K, Amos W (1997) Behavioral, ecological and molecular genetic analyses of reproductive strategies in the Amazonian dart-poison frog, *Dendrobates ventrimaculatus*. Behav Ecol 8:260–267

Tamura K, Dudley J, Nei M, Kumar S (2007) *MEGA4*: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

This P, Jung A, Boccacci P, Borrego J, Botta R, Costantini L, Crespan M, Dangl GS, Eisenheld C, Ferreira-Monteiro F, Grando S, Ibáñez J, Lacombe T, Laucou V, Magalhães R, Meredith CP, Milani N, Peterlunger E, Regner F, Zulini L, Maul E (2004) Development of a standard set of microsatellite reference alleles for identification of grape cultivars. Theor Appl Genet 109:1448–1458

This P, Lacombe T, Thomas MR (2006) Historical origins and genetic diversity of wine grapes. Trends Genet 22(9):511–519

Vouillamoz JF, Grando MS (2006) Genealogy of wine grape cultivars: 'Pinot' is related to 'Syrah'. Heredity 97:102–110

Vouillamoz JF, Monaco A, Costantini L, Stefanini M, Scienza A, Grando MS (2007) The parentage of 'Sangiovese', the most important Italian wine grape. Vitis 46(1):19–22

Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Mol Ecol Notes 10:249–256

Wright S (1951) The genetical structure of populations. Ann Eugen 15:323–354

Zohary D, Hopf M (2000) Domestication of plants in the old world, 3rd edn. Oxford University Press, Oxford