CrossMark

ORIGINAL ARTICLE

High-throughput 18K SNP array to assess genetic variability of the main grapevine cultivars from Sicily

Francesco Mercati[1,2] · Gabriella De Lorenzis[3] · Lucio Brancadoro[3] · Antonio Lupini[1] ·
Maria Rosa Abenavoli[1] · Maria Gabriella Barbagallo[4] · Rosario Di Lorenzo[4] ·
Attilio Scienza[3] · Francesco Sunseri[1]

Abstract The viticulture of Sicily, for its vocation, is one of the most important and ancient forms in Italy. Autochthonous grapevine cultivars, many of which known throughout the world, have always been cultivated in the island from many centuries. With the aim to preserve this large grapevine diversity, previous studies have already started to assess the genetic variability among the Sicilian cultivars by using morphological and microsatellite markers. In this study, simple sequence repeat (SSR) were utilized to verify the true-to-typeness of a large clone collection (101) belonging to 21 biotypes of the most 10 cultivated Sicilian cultivars. Afterwards, 42 Organization Internationale de la Vigne et du Vin (OIV) descriptors and a high-throughput single nucleotide polymorphism (SNP) genotyping array (Vitis18kSNP) were applied to assess genetic variability among cultivars and biotypes of the same cultivar. Ampelographic traits and high-throughput SNP genotyping platforms provided an accuracy estimation of genetic diversity in the Sicilian germplasm, showing the relationships among cultivars by cluster and multivariate analyses. The large SNP panel defined sub-clusters unable to discern among biotypes, previously classified by ampelographic analysis, belonging to each cultivar. These results suggested that a very large number of SNP did not cover the genome regions harboring few morphological traits. Genetic structure of the collection revealed a clear optimum number of groups for $K = 3$, clustering in the same group a significant portion of family-related genotypes. Parentage analysis highlighted significant relationships among Sicilian grape cultivars and Sangiovese, as already reported, but also the first evidences of the relationships between Nero d'Avola and both Inzolia and Catarratto. Finally, a small panel of highly informative markers (12 SNPs) allowed us to isolate a private profile for each Sicilian cultivar, providing a new tool for cultivar identification.

Keywords Biotype identification · Single nucleotide polymorphism (SNP) · Parentage analysis · *Vitis vinifera* L.

Francesco Mercati and Gabriella De Lorenzis contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s11295-016-1021-z) contains supplementary material, which is available to authorized users.

✉ Francesco Sunseri
  francesco.sunseri@unirc.it

[1] Dipartimento AGRARIA, Università Mediterranea di Reggio Calabria, Località Feo di Vito, 89121 Reggio Calabria, Italy

[2] Istituto di Bioscienze e Biorisorse (IBBR), Consiglio Nazionale delle Ricerche (CNR), Corso Calatafimi, 414, 90120 Palermo, Italy

[3] Dipartimento di Scienze Agrarie ed Ambientali, Università degli Studi di Milano, Via Celoria, 2, 20133 Milano, Italy

[4] Dipartimento di Scienze Agrarie e Forestali, Università degli Studi di Palermo, Viale delle Scienze, 11, 90128 Palermo, Italy

Introduction

Vitis vinifera L., one of the most widely cultivated species of agricultural and economical interest (Vivier and Pretorius 2002), is of Indo-European origin and its distribution area extends from Central Asia to the Mediterranean Basin (Zohary et al. 2000). Over the centuries, different events (such as domestication process and outbreeding) have had a significant influence on the large genetic pool of grapevine cultivars around the word, leading grapevine to be one of the most heterozygous species, carrying in its genome deletions,

insertions, inversions, and single nucleotide polymorphisms (This et al. 2006; Jaillon et al. 2007; Velasco et al. 2007). The accumulation of casual mutations and natural or artificial crossing have been the driver of grapevine evolution since its domestication (This et al. 2006; Forni 2012). The mutations could happen in shoot apical meristematic layers with different fate: (1) they are preserved by asexual propagation when occur in the somatic cell lines, or (2) they are inherited to the progenies by sexual reproduction when they arose in the germinal cell lines (D'Amato 1997). Because the grapevine propagation is done mainly through cuttings, somatic mutagenic events can cumulate over time, leading to determine genetic variation and the creation of new cultivars and clones (This et al. 2006). The induced polymorphisms are often detectable into the genome as single nucleotide polymorphism (SNP) and insertion/deletion (INDEL), and several examples were already reported. The *Gret1* retrotransposon insertion in the *VvMybA1* promoter region is one of the induced variability example detected in the Pinot family cultivars, differing for the berry color (Kobayashi et al. 2004; Yakushiji et al. 2006; Vezzulli et al. 2012). The Chardonnay musqué clone 44–60 Dijon, differing for a SNP in the candidate gene for muscat flavor from the mother clone (Emanuelli et al. 2010), and Carignan, where clones showing different cluster shape harbor the insertion of *Hatvine1-rrm* transposable element in the *VvTFL1A* promoter, confirmed these phenomena (Fernandez et al. 2010). Due to this wide genetic variability, clonal selection is the most common breeding method to improve grapevine by exploiting intra-varietal genetic diversity and identifying high-performant clones in a specific environment and their consequent official registration, propagation, and distribution to viniculture market.

With the aim to detect polymorphisms and distinguish among grapevine cultivars, different molecular markers such as RAPD (Moreno et al. 1995), AFLP (Cervera et al. 1998), SSR (Bowers et al. 1996; Sefc et al. 1999), ISSR (Moreno et al. 1998), S-SAP (D'Onofrio et al. 2010), REMAP (Castro et al. 2012), and SAMPL (Cretazzo et al. 2010) were used. The microsatellites or simple sequence repeats (SSRs) are highly reproducible and informative for their codominant and multiallelic fashion, playing a predominant role to evaluate genetic diversity in several plant crops, such as maize (Reif et al. 2006), rice (Thomson et al. 2007), wheat (Laidò et al. 2013), peach (Dirlewanger et al. 2002), olive (Cipriani et al. 2002), and citrus (Barkley et al. 2006). Also, in grapevine, genotyping is mainly based on SSR, which have been useful for cultivar identification, finding relationships among cultivars, synonyms, homonyms (Cipriani et al. 2010; Laucou et al. 2011; Emanuelli et al. 2013), and parentage analysis (Lacombe et al. 2013). However, SSR showed their limit to help clonal selection and identification, being not always able to discriminate among clones/biotypes of the same cultivar (González Techera et al. 2004; Pelsy et al. 2010). Thus, based

also on SSR time consuming and costly detection, other markers should be developed to detect somatic mutations and clonal variations.

The advent of high-throughput next-generation sequencing (NGS) technologies with the possibility to sequence entire genomes more efficiently allowed to obtain large-scale SNP identification for crops and the onset of efficiency SNP genotyping platform (Schmid et al. 2003; Dereeper et al. 2011; Chagné et al. 2012; Ganal et al. 2012; Peace et al. 2012; Verde et al. 2012; Gardner et al. 2014; Yu et al. 2014; Tayeh et al. 2015; Jiang et al. 2016; Melo et al. 2016). The abundance in the genome and the ability to identify polymorphism due to variation at single base level based on bi-allelic nature are the main advantages of SNP markers. In addition, the high-throughput multiplexed SNP assay represents a very useful tool to evaluate genome-wide allelic variation for genetic diversity, population structure, and parentage analysis in crops. Although SNP polymorphism information content (PIC) is lower compared to SSR, the high number of identifiable SNP in the genome and their reproducibility make them ideal for developing panels of markers useful for genetic variation and cultivar identification (Myles et al. 2011; Sim et al. 2012; Mercati et al. 2015; Winfield et al. 2015; Kurokawa et al. 2016).

The grapevine genome sequence has been available since 2007 (Jaillon et al. 2007; Velasco et al. 2007), and large-scale SNP discovery and genotyping have been reported (Lijavetzky et al. 2007; Pindo et al. 2008), leading to identify more four hundreds of thousands SNPs, validating a subset on a 9K genotyping array (Myles et al. 2010, 2011). Exploiting SNP frequency and their discrimination power, informative SNPs set for cultivar identification (Cabezas et al. 2011), and clonal variation studies were also identified (Carrier et al. 2012). More recently, in the frame of a large-scale grape genome resequencing (GrapeReSeq Consortium), a high-throughput genotyping 18K SNP chip was developed (Le Paslier et al. 2013).

Among the European grapevine regions, Sicily is devoted to the viticulture from ancient time and characterized by a rich ampelographic platform including cultivars both of local, national, and international interest. This genetic diversity has been probably preserved by on-field grafting traditional technique, in contrast to the widely usage of pre-grafted and certified cultivars coming from a narrow genetic pool, adopted in other European grapevine regions. In the past decade, more than 3000 accessions around the island were collected and included in a clonal selection program. Based on morphological traits (48 Organization Internationale de la Vigne et du Vin (OIV) descriptors utilized in the GrapeGen06 European project; www.eu-vitis.de/index.php) (Maul et al. 2012), more plants from different biotypes belonging to each cultivar were selected (Maitti et al. 2009). In viticulture, the term biotype refers to morphological variations within a cultivar. Single

plants or different clonal lines are classified in the same biotype if sharing similar phenotypic expressions regarding morphological traits, such as bunches and/or leaves, slightly different from the most frequent phenotypic traits of the cultivar. This morphological variability is usually addressed to the long-time cultivation of variety in a specific geographical area (Campostrini et al. 1995). Those morphological traits have an effect on the qualitative characteristics of grapes and musts that arise. A typical example is the different oenological aptitude to wine-making of Pinot noir clones. Indeed, the biotypes with large berries are used to produce sparkling wine, while that with small berries are used for red wine, as reported in the catalogue of vines grown in France (http://plantgrape. plantnet-project.org). In the context of the clonal selection program in Sicily, the genetic variability among the main cultivars was assessed based also on selected SSR panel (Carimi et al. 2010, 2011; De Lorenzis et al. 2014).

In the present study, a grapevine core collection from Sicily was characterized by high-throughput Vitis18kSNP genotyping array to assess the genetic relationships among cultivars and to discriminate among biotypes of the same cultivar. Our results well distinguished the Sicilian cultivars, providing additional information about their genetic relationships by parentage analysis. In such cases, SNP genotyping appears also able to differentiate among biotypes of the same cultivar. Furthermore, the present study allowed us to isolate a small panel (12) of highly informative SNPs that may become a rapid diagnostic tool for Sicilian cultivar identification.

Material and methods

Plant material

A panel of 101 samples from 21 biotypes belonging to 10 Sicilian grapevine cultivars, included in the grapevine collection of Regione Sicilia (Marsala, Italy), was taken into account. Based on the ampelographic traits recorded for each cultivar in the experimental field, from 1 to 3 biotypes per cultivar were considered (Table 1 and supplementary file, Table S1). For each cultivar, clones already registered at the National Register of Grapevine cultivars (http://catalogoviti. politicheagricole.it/) were included in the panel. Catarratto Bianco Comune and Catarratto Bianco Lucido are two varieties registered at the National catalogue as distinct cultivars, even though they showed identical genetic profile at 11 SSR loci (Crespan et al. 2008). The first one corresponds to Catarratto biotype A and B and the second one to biotype C. Pinot Noir and Sangiovese were included in the analysis as international and national reference varieties, respectively.

Ampelographic analysis and SSR genotyping

Ten plants were cloned from each of the 21 biotypes (belonging 10 Sicilian cultivars) and were utilized for the ampelographic analysis to assess intra-cultivar variability (Table 1). Forty-two ampelographic traits (supplementary file, Table S2), related to young shoot, shoot, young and mature leaf, woody shoot, bunch, and berry, were recorded as specified by the OIV (http://www.oiv.int/) during spring-summer seasons 2011 and 2012. The observations were carried out in different times during the vegetative seasons, such as at flowering for OIV 1 or between berry set and *vèraison* for OIV 65, as reported in the second edition of the OIV Descriptor List for Grape Varieties and *Vitis* Species (http://www.oiv.int/en/technical-standards-and-documents/description-grape-varieties/oiv-descriptor-list-grape-varieties-and-vitis-species-2nd-edition).

A detailed description of each trait and its expression is reported at http://www.eu-vitis.de/docs/descriptors/mcpd/WP2-DESCRIPTORS-v4.pdf. Finally, using the ggplot2 R

Table 1 List of the major Sicilian *V. vinifera* cultivars analyzed by Vitis18kSNP array

Cultivars	Biotypes	Number of accessions	Total
Carricante	A	5	
	B	4	9
Catarratto[a]	A	8	
	B	4	
	C	5	17
Frappato	A	5	
	B	5	10
Grecanico	A	6	
	B	7	13
Grillo	A	2	
	B	2	4
Inzolia (synonym Ansonica)	A	4	
	B	3	7
Nerello Cappuccio	A	3	3
Nero d'Avola (synonym Calabrese)	A	9	
	B	13	
	C	7	29
Perricone	A	3	
	B	3	6
Zibibbo	A	1	
	B	2	3
Total	21	101	101

[a] Biotypes A and B are two biotypes of Catarratto Bianco Comune and biotype C is biotype of Catarratto Bianco Lucido, two varieties registered at the National Register of Grapevine cultivar (http://catalogoviti. politicheagricole.it/home.php), but showing the same microsatellite allele profile (Crespan et al. 2008)

package (https://cran.r-project.org/web/packages/ggplot2/index.html), a heatmap describing the variation of OIV descriptors were set up. Each descriptor was recorded on a 1–9 scale, and the different colors and gradients were associated to the scale and combination for each category.

All 101 clones and 2 reference cultivars were analyzed by using the 9 SSR (VrZag62, VrZag79, VVMD5, VVMD7, VVMD25, VVMD27, VVMD28, VVMD32, VVS2) (Thomas and Scott 1993; Bowers et al. 1996, 1999a; Sefc et al. 1999), suggested as a standard set for grapevine genotyping in the frame of the GrapeGen06 European project, to assess the common genetic background of biotypes belonging the same cultivar. Genomic DNA was extracted from 0.1 g of young leaves tissue (1–2 cm of diameter), using the QiagenDNeasy Plant Mini Kit (Qiagen, Hilden, Germany). DNA quality (260/230 and 260/280 ratios) and concentration was checked by NanoDrop Spectrophotometer (Thermo Scientific, Waltham, MA). Multiplexed amplification reactions were performed in 25 μl final volume reaction mixture as described in De Lorenzis et al. (2013). The amplification products were solved on ABI PRISM 310 Genetic Analyser (Applied Biosystems by Life Technologies, Foster City, USA), and the alleles were sized by GENEMAPPER 4.0 (Applied Biosystems by Life Technologies). Pinot Noir and Sangiovese were included as reference varieties for the allele standardization.

SNP array analysis and reproducibility

The genotyping of the whole studied panel (101 Sicilian grapevine genotypes and 2 reference varieties) was carried out using the custom Vitis18kSNP array (Illumina Inc., San Diego, CA), designed by GrapeReSeq Consortium, which assays 18,071 SNPs (Le Paslier et al. 2013). DNAs, extracted as reported above, were delivered to Fondazione Edmund Mach (San Michele all'Adige, Trento, Italy) and to TraitGenetics GmbH (Gatersleben, Salzlandkreis, Germany) for genotyping. Two hundred nanograms of genomic DNA were used as template for the reaction, following the manufacturer's instructions (Illumina Inc.). Because of genotypes were processed in two different service platforms (Fondazione Edmund Much and TraitGenetics GmbH), one sample per each biotype from eight out of ten Sicilian cultivars was genotyped twice (one per each laboratory) starting from two independent DNA extractions, to measure the reproducibility of Vitis18kSNP assay. The differences among duplicated SNP profiles were evaluated, and cluster analysis was performed to establish the threshold reproducibility of our system, according to unweighted pair-group method with arithmetical average (UPGMA) algorithm by Molecular Evolutionary Genetics Analysis (MEGA) version 5 (Tamura et al. 2011). Percentage of genetic similarity among replicates of each biotype was determined by cluster analysis, and the lowest value of

similarity that grouped the replicates of a same biotype was used to determine the threshold value of reproducibility (Okitsu et al. 2013).

The reproducibility value was high, determining the threshold above 99 %. Among cultivars, Inzolia showed the highest percentage of different loci (0.36 %, 41 SNP loci), while Carricante the lowest ones (0.12 %, 14 SNP loci; Table S3). The results confirmed the stability of data produced.

SNP data processing and structure population analysis

SNP row data were scored with Genotyping Module 1.9.4 of the GenomeStudio Data Analysis V2011.1 software (Illumina Inc.). The dataset was filtered and standardized utilizing the following criteria: (i) samples with low SNP call quality ($p50GC < 0.54$) were removed; (ii) SNPs with a GenTrain score higher than 0.6 were retained; (iii) the monomorphic SNPs were detached; (iv) SNPs with a number of non-calls (NCs) higher than 20 % were deleted. For all the analyses SNPs with minor allele frequency (MAF) < 0.05, and missing rate > 0.20 were removed.

The main genetic parameters to analyze population genetic diversity, including observed (H_o) and expected heterozygosity (H_e) (Nei 1973), the MAF, and the inbreeding coefficient (F), were carried out by PEAS 1.0 software (Xu et al. 2010).

The SNP dataset was utilized to investigate the genetic relationships among biotypes of ten major Sicilian cultivars by both principal coordinates analysis (PCoA) and cluster analysis. PCoA was performed using SNPrelate, an R package for large-scale calculations (Zheng et al. 2012). A linkage disequilibrium (LD)-based pruned SNP set was first choice based on 0.2 LD thresholds to avoid a large SNP clusters effect. LD-based pruned SNP set was utilized for PCoA analysis by using the snpgdsPCA function in SNPrelate. A phylogenetic tree was designed by the UPGMA method, implemented in the MEGA 5.0 software (Tamura et al. 2011). The bootstrap analysis was performed based on 100 resampling.

Population structure analysis (Pritchard et al. 2000), a Bayesian approach to inferring the correlation between genotypes based on admixture model, was performed using fastSTRUCTURE package (Raj et al. 2014). Individuals were assigned to K population/genetic clusters based on their multilocus profile. The genetic clusters were assembled to minimize intra-cluster LD, and the proportion of membership for each genotype was estimated. The admixture model without the prior population information was employed. K values from 1 to 10 were tested, and the most likely K was chosen running the algorithm for multiple choices of K (Raj et al. 2014). For each run, the initial burn-in period was set to 50,000 iterations with 500,000 Markov chain Monte Carlo (MCMC) replications. The admixture proportions estimating the most likely K were viewing by DISTRUCT software (Rosenberg 2004).

The identification of highly informative SNPs for varietal identification was also carried out. The R package Genome Association and Prediction Integrated Tool (GAPIT) (Lipka et al. 2012) was used with default parameters to identify private SNP profiles related to the cultivars analyzed. For GAPIT analysis, 12 categories (one for each cultivar) were assigned, based on SSR profiles and ampelographic data. A mixed linear model (MLM) approach (Yu et al. 2006; Zhang et al. 2010) was adopted and a kinship (k) matrix was calculated. The Benjamini-Hochberg procedure (Benjamini and Hochberg 1995) adjusted for the multiple testing problems by controlling the false discovery rate (FDR) at 0.05 and the missing data were treated by major allele substitution. To identify specific SNP profiles related to the germplasm studied, each SNP was tested in turn, using an F test (H0: no association between the SNP profile and cultivar) and P values were obtained (Lipka et al. 2012). The SNPs selected were randomly verified through Sanger sequencing method following standard protocol to confirm the marker profiles identified by GAPIT. The PCR reaction was carried out in a 20 μl volume containing 50 ng of genomic DNA, 1× supplied PCR buffer (Promega), 0.2 mM of each dNTP (Roche), 0.2 unit of Taq DNA polymerase (Promega), 0.20 μM of specific primers pair. PCR reactions were performed under the following cycle program: 94 °C (5 min), then 30 cycles at 94 °C (30 s)/60 °C (30 s)/72 °C (30 s), and a final extension at 72 °C for 10 min. Subsequently, the fragments were solved on ABI PRISM 310 Genetic Analyser (Applied Biosystems by Life Technologies, Foster City, USA). The primer pairs are listed in supplementary file (Table S4).

Parentage analysis

The identity-by-descent (IBD) index (the probability that two genotypes are descended from single ancestral genotype and not identical by chance) was carried out on each pair of genotypes by PLINK 1.07 software (Purcell et al. 2007) to infer relationships among non-redundant individuals. The filtered SNP dataset was used, and the most frequent SNP profile for each cultivar was chosen for parentage analysis. MAF and r^2 of LD were set on 0.01 and 0.05 values, respectively. Four parameters were taken into account, Z0 (probability of sharing 0 IBD allele identical-by-descent), Z1 (probability to share 1 IBD allele), Z2 (probability to share 2 IBD alleles), and PI-HAT [the relatedness measure measured as PI-HAT = P (IBD = 2) + 0.5 × P (IBD = 1)]. In parent-offspring relationships, Z0 and Z2 are expected to be 0 and Z1 to be 1, while in second-degree pairs, Z0 and Z1 are expected to be 0.5 and Z2 to be 0. Therefore, pairs of genotypes showing a PI-HAT value similar to 0.5 are related by first-degree or closer relationships. Starting from the PLINK results, a circular plot was developed, reporting the first- and second-degree relationships among varieties.

Results

Ampelographic analysis

Morphological traits were scored twice during spring-summer seasons 2011 and 2012, recording 42 out of 48 OIV descriptors suggested by the European GrapeGen06 project (Maul et al. 2012) from ten cloned plants of each biotype. A detailed description of the 42 OIV descriptors, reporting the main traits discriminating among biotypes, is included in the supplementary file (Table S2). For an easy-view of the ampelographic results, a heatmap was provided, representing the expression level of each OIV descriptor in each biotype (Fig. 1). The biotypes clustered according to their own cultivar and the differences among biotypes of the same cultivar were displayed as well, even though those differences were addressed to few descriptors. The descriptors OIV 6, 51, 65,
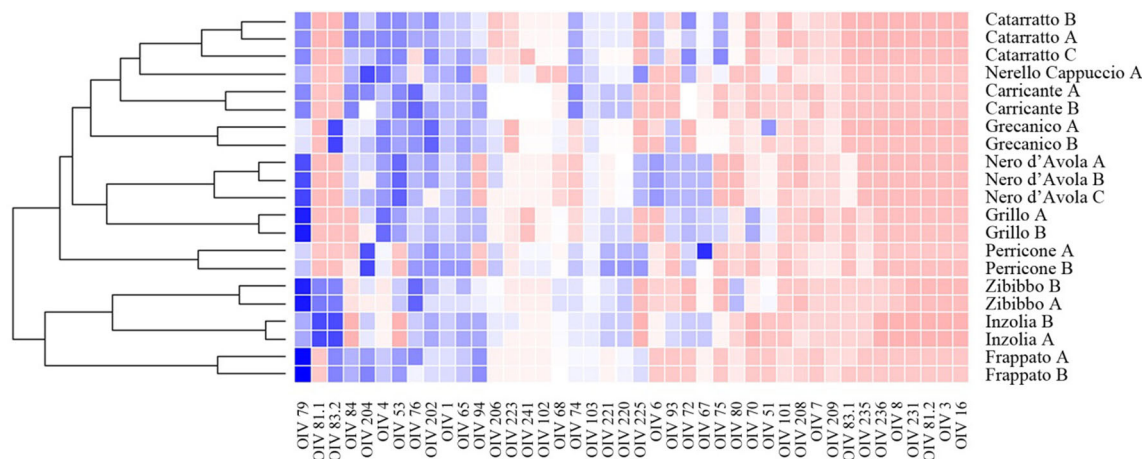

Fig. 1 Forty-two OIV descriptors recorded for each biotype of ten Sicilian grapevine cultivars. The expression level of OIV descriptors has been represented by a heatmap. *Different colors and gradients* represent the categories for each descriptor reported in detail in Table S2

67, 68, 75, 76, 83–1, 84, 93, 94, 101, 202, 204, 206, 208, 220, 221, 223, and 241 showed differences among cultivars that can be clearly distinguished, while the discrimination among biotypes of the same cultivars was much less evident (Fig. 1, Table S2). The descriptors showing the highest differences in their expression levels among the biotypes of the same cultivar were OIV 202 (length of bunch), with values ranging from 3 (short) and 9 (very long), and OIV 204 (density of bunch), with values ranging from 3 (loose) to 9 (very dense) overall the samples. Catarratto showed the highest number of discriminant traits, indeed 14 OIV descriptors were able to discriminate biotype C from A and B. The other cultivars showed a less number of different ampelographic traits, ranging from 1 (Grecanico) to 4 (Inzolia and Nero d'Avola), mainly related to the bunch morphology (Fig. 1).

SSR-based true-to-typeness cultivar determination

Nine SSR markers selected in the frame of European GrapeGen06 project were utilized to establish the genetic profile of each biotype belonging to their own cultivar. Twelve SSR profiles were obtained, one for each cultivar, including the two reference cultivars, Pinot Noir and Sangiovese (supplementary file, Table S5). Biotypes of the same cultivars held the same SSR profile, confirming their common background. The true-to-typeness of each cultivar was determined by a comparison with already public standardized SSR profiles (Italian Vitis Database (IVD), http://www.vitisdb.it; Vitis International Variety Catalogue (VIVC), http://www.vivc.de/index.php).

SNP analysis and genetic relationship assessment among cultivars and biotypes

The genetic relationships among the main Sicilian grapevine cultivars and the intra-varietal genetic variation were deeper investigated by using the Vitis18kSNP array, a high-throughput genotyping system. After the SNP dataset inspection, 554 loci (3 %) did not amplify among all genotypes and about 14,794 loci (82 %) showed GT score higher than 0.6 (Table 2). The final dataset resulted in 14,755 out of 18,071 loci, after removing the SNPs with a number of not calling (NC) higher than 20 %. The number of polymorphic loci (11,411) was about 77 % of the final SNP panel (Table 2). The expected and observed heterozygosity values were quite similar (0.284 and 0.313 for H_e and H_o, respectively) with a mean inbreeding coefficient of −0.102 (Table 2). The overall value of MAF was 0.210 and 3643 out of 11,411 SNP loci (about 32 %) showed a MAF value lower than 0.100.

The SNP profiles of samples belonging to their own cultivar were nearly identical, although the SNP divergences among plants of the same cultivar ranged from eight (Nerello Cappuccio) to 247 (Catarratto) loci (mean value of

Table 2 Summary statistics of genetic variation obtained by Vitis18kSNP array in 101 Sicilian grapevine clones belonging to 21 biotypes and 10 cultivars

Parameters	Values
N^a	101
Total no. of loci	18,071
100 % failed loci	554
Higher than 0.6 GT	14,794
No. of used loci	14,755
No. of polymorphic loci	11,411
H_e	0.284
H_o	0.313
MAF	0.210
F	−0.102

GT GenTrain score, H_e expected heterozygosity, H_o observed heterozygosity, *MAF* minor allele frequency, *F* inbreeding coefficient

[a] Sample size

52.5). The list of polymorphic loci for each cultivar was reported in the supplementary file (Table S6). Almost the whole cases of polymorphism (99.9 %) were due to the changes from homozygous to heterozygous loci. Only in few cases (0.1 %), the polymorphism was determined by the change from a homozygous to the other one, probably due to the natural nucleotide variation that allowed the fixation or loss of a new mutation in specific loci. The position 14879091 in chromosome 7 was the polymorphic locus shared among biotypes of eight analyzed cultivars. The major number (4) of polymorphic SNP loci were shared from the pair Frappato - Nero d'Avola (2:16685674; 1:9046299; 5:10489883; 11:6480292), followed by Catarratto - Perricone, Frappato - Grecanico, and Perricone – Zibibbo pair, which had three common polymorphic SNP loci.

Multivariate and population structure analyses

Multivariate PCoA and cluster analysis were used to investigate the genetic distances among cultivars and biotypes by using 11,411 polymorphic SNP loci. PCoA, a scattered plot reporting the first two coordinates and describing all cultivars analyzed, is showed in Fig. 2a. The main two coordinated explained 32.52 and 21.31 % of total variability, respectively. Most of the samples (96 %) were grouped based on their own biotype and/or cultivar, with the exception of Catarratto, which showed the highest genetic polymorphism among genotypes. Cluster analysis assigned properly all samples (100 %) to their own cultivar (Fig. 2b), likewise PCoA, showing Grecanico and Nero d'Avola as the two most distant cultivars based on SNP genetic diversity. Cluster analysis highlighted also that discrimination among biotypes of the same cultivar appeared difficult, indeed the sub-clusters for each cluster included different biotypes. The bootstraps
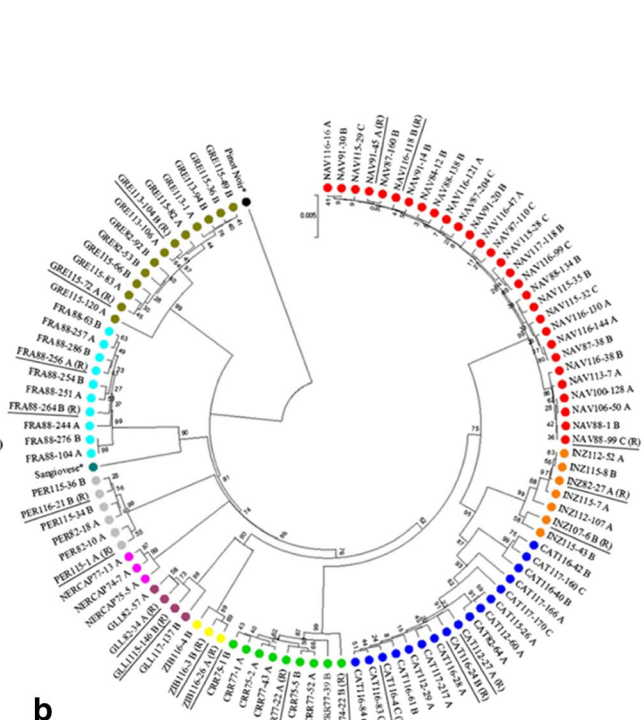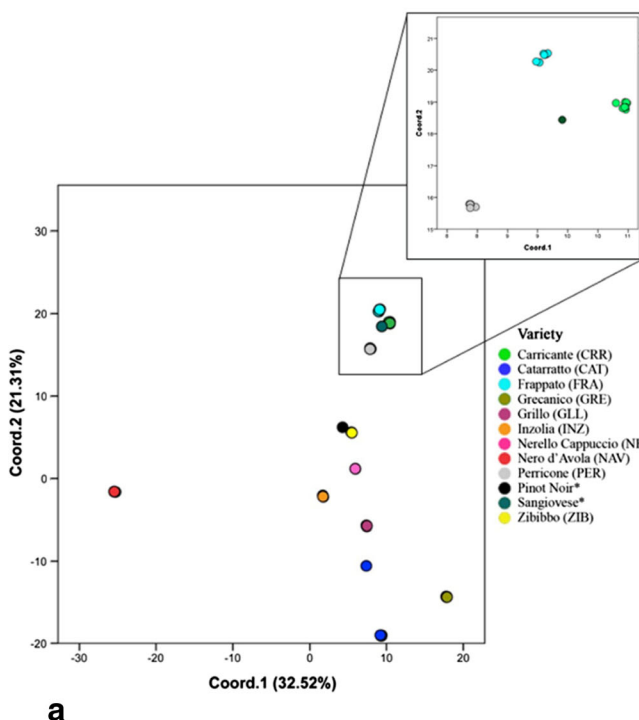
Fig. 2 Genetic relationships among 101 Sicilian grapevine clones, arranged in 10 cultivars and 21 biotypes, and two reference cultivars, obtained by 11,411 polymorphic SNPs. **a** Principal coordinates analysis (PCoA) based on Vitis18kSNP. In the *brackets* the percentage of variability explained by each coordinate is reported. A zoom in has been provide in order to highlight the differences among clones of the same cultiva. **b** Dendrogram generated using UPGMA method. Bootstrap values are shown as percentage. **a–c** different biotypes of the same cultivar; *reference cultivar

among cluster, ranging from 95 to 99 %, should avoid misclassifications.

The allelic profiles were used in the model-based clustering method implemented by fastSTRUCTURE software to ascertain the likely number of genetic groups (K) within the Sicilian germplasm collection. The algorithm for multiple choices of K revealed a clear optimum for $K=3$. The genetic structure of analyzed panel is shown in Fig. 3. Eighty out of 101 clones (about 76 %) were assigned to a cluster at $K=3$ by using a >80 % threshold for group classification. Interestingly, none biotypes/genotypes were misclassified based on their own cultivar. Nearly the totality of cultivars showed 100 % of membership for one group. Indeed, Nero d'Avola, showing a private genetic structure, represented as the green pool (Fig. 3), Carricante, Frappato and Perricone together with Nerello Cappuccio and Zibibbo were members of the purple group (Fig. 3), while Grillo, Catarratto and Grecanico represented the third group (blue; Fig. 3). The analysis was not able to assign Inzolia to a specific genetic pool, since the percentages of membership for groups 1, 2, and 3 were 38, 24 and 38 %, respectively.

Different default criteria, as high frequency of genotyping success (missing rate < 0.20) and MAF > 5 %, were adopted to select an informative SNPs dataset (7235), starting from 11,
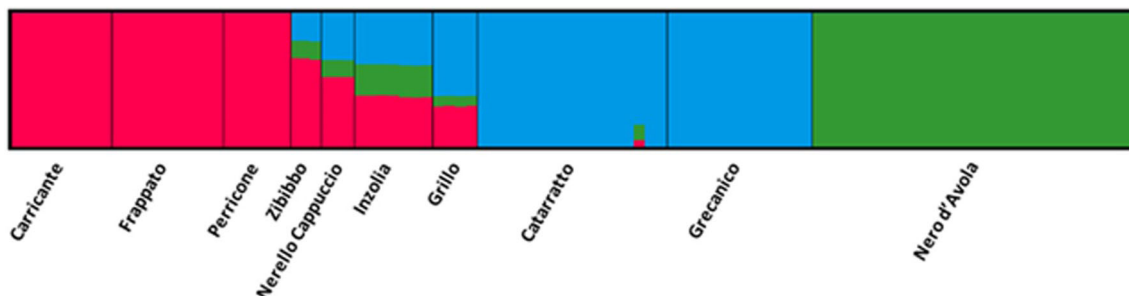

Fig. 3 Admixture proportions of Sicilian grapevine cultivars, as estimated by fastSTRUCTURE ($K = 3$) based on 11,411 polymorphic SNPs from the Vitis18kSNP. The results are grouped by cultivars of origin for each individual. *Each vertical bar* represents a sample (101 clones belonging 10 Sicilian cultivars). The *color proportion for each bar* represents the posterior probability of assignment of each individual to one of three groups of genetic similarity. The range of assignment probability varies from 0 to 100 %

411 polymorphic SNPs used in the genetic diversity analysis. The selected panel was used to identify putative private profiles related to each cultivar. Twelve categories, one for each cultivar (ten Sicilian and two reference grapevine cultivars), were assigned based on the SSR profiles and ampelographic analysis. The identification of specific markers profile associated to the 12 categories was performed using a MLM, controlling the relatedness based on kinship values. The MLM was applied to study the links between marker profiles and cultivars, since it improves the ability to detect phenotype-genotype associations in presence of population stratification and multiple levels of relatedness, increasing the statistical significance of the analysis. The kinship matrix was calculated based on the percentage of shared alleles, displaying the clustering of cultivars and the dissimilarity among genotypes (Fig. 4). This analysis confirmed private genetic clusters for each investigated cultivars as already suggested by the previous analyses. Indeed, the samples belonging to the same cultivar clustered in a common branch, of which Grecanico and Nero d'Avola are the two varieties most distant (Fig. 4). Although a large set of genetic markers that provide good

coverage of whole genome was used, the kinship matrix, as well as the cluster analysis, did not allow us to discriminate among biotypes of the same cultivar. However, the chosen approach allowed us to identify a set of highly informative markers (12 SNPs, in both coding and non-coding regions) that, through their profiles combination (Table 3), can discriminate all the cultivars included in the present study, except Catarratto (Table 3). Indeed, this cultivar showed three different profiles, one of which is prevalent (14 out of 17 samples); however, all profiles were able to discriminate this cultivar from the others (Table 3). To verify the private profiles identified using the high-throughput genotyping system and MLM approach, the isolated SNPs were randomly tested on the same cultivars studied and validated through Sanger method. The analysis of sequences around the most informative SNP confirmed the private profiles belonging to each cultivar.

Parentage analysis

SNP dataset and the probability to have IBD alleles were probed to investigate the parentage among cultivars, assigning
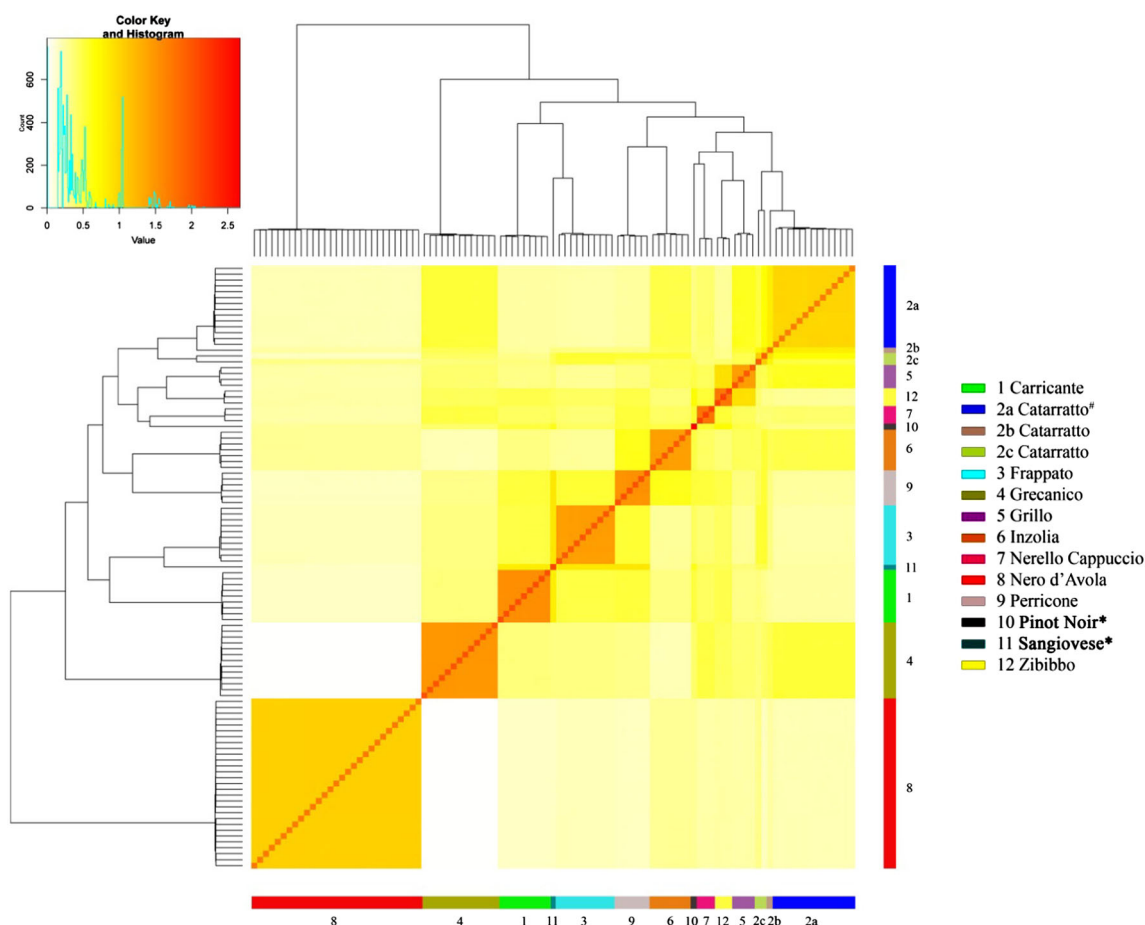

Fig. 4 Kinship analysis for ten Sicilian and two international-national grapevine cultivars, visualized through a heatmap based on 11,411 polymorphic SNPs from the Vitis18kSNP. Color gradient displays the dissimilarity among genotypes: *red* indicates the most similar clones, while *white* shows the lowest genetic similarity. The cultivars are indicated with colors used in the PCoA and clustering analysis. Reference varieties (* and *bold*); #Catarratto prevalent profile

Table 3 SNP private profiles of 10 Sicilian grapevine cultivars at 12 highly informative loci out of 18,071, able to discriminate among the analyzed varieties

	SNP											
Cultivars	rs13625	rs13626	rs11820	rs10957	rs9788	rs5476	rs5488	rs5508	rs10557	rs7963	rs4413	rs5516
Carricante (9/9)	C	G	C	Y	M	G	A	G	G	K	Y	R
Catarratto[a] (14/17)	C	G	Y	T	M	K	R	R	R	K	Y	R
Catarratto (2/17)	Y	R	Y	Y	M	K	R	R	R	K	Y	R
Catarratto (1/17)	C	G	Y	Y	M	K	R	R	R	K	Y	R
Frappato (10/10)	Y	R	Y	Y	M	K	R	R	A	G	Y	R
Grecanico (13/13)	C	G	Y	Y	M	K	R	R	A	K	Y	R
Grillo (4/4)	Y	R	Y	Y	M	K	R	R	R	K	Y	G
Inzolia (7/7)	Y	R	C	Y	M	T	G	A	A	K	Y	G
Nerello Cappuccio (3/3)	Y	R	C	C	M	T	G	A	A	T	Y	G
Nero d'Avola (29/29)	Y	R	C	Y	C	T	G	A	A	T	Y	G
Perricone (6/6)	T	A	C	C	M	T	G	A	A	T	C	G
Zibibbo (3/3)	T	A	C	C	C	T	G	A	A	T	C	G
Pinot Noir[b]	Y	R	C	C	C	T	G	A	A	T	C	G
Sangiovese[b]	T	A	C	C	C	K	R	R	A	K	C	G

In brackets, the number of samples classified in the private profile

[a] Catarratto prevailing profile

[b] Reference cultivars

the properly relationship category, such as parent-offspring and second degree. The most informative relationships (first and second degree) were displayed in the circular plot (Fig. 5). A complete list of relationship categories per each pairwise genotype is also recorded in Table S7. Ten out of 12 cultivars resulted to be related to other cultivars included in the panel, for a total of 7 relationships, five of them classified as PO

(parent-offspring), showing Z0, Z1, Z2, and PI-HAT values similar to theoretical values (0, 1, 0, and 0.5, respectively), and the two pairs showed a second-degree relationship, having relatedness values similar to theoretical values (0.5, 0.5, 0, 0.25).

The cultivar Catarratto showed the highest number of relationships within the analyzed panel, two PO (with

Fig. 5 Circular plot displaying the first- and second-degree relationships among ten Sicilian grapevine cultivars based on polymorphic SNPs from the Vitis18kSNP. The cultivars are represented as *circles* sited at the corners of a 12-side geometric shape; the *black lines* represent the architecture of the geometric shape and do not correspond to any parentage relationship among cultivars. *Filled circles*: Sicilian grapevine cultivars; *empty circles*: reference varieties; *green lines*: first-degree relationships; *blue lines*: second-degree relationships
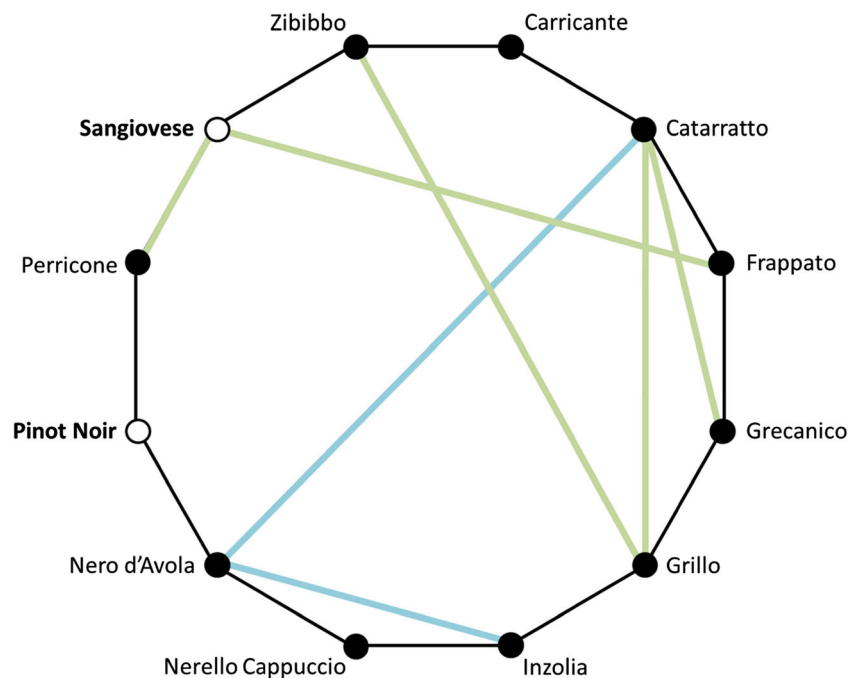
Grecanico and Grillo) and one second degree (with Nero d'Avola) (Fig. 5 and supplementary file, Table S7). Among Sicilian cultivars, Carricante and Nerello Cappuccio did not show any relationships with the other cultivars. As expected, Pinot Noir did not show parentage relationships with all the other cultivars, while Sangiovese showed two PO (with Frappato and Perricone) relationships with the Sicilian cultivars. Catarratto, Inzolia, and Nero d'Avola linked each other by a second-degree relationship, even though the Catarratto-Inzolia pair showed relatedness values rather deviated from theoretical values (supplementary file, Table S7). Finally, the pairwise Carricante-Sangiovese showed the empiric values in between theoretical values for second-degree relationship and unrelated genotypes (supplementary file, Table S7).

Discussion

The genetic variability among 21 biotypes belonging to 10 major Sicilian autochthonous grapevine cultivars (101 clones) and their relationships were investigated by either 42 ampelographic OIV descriptors, 9 SSR or 18,701 SNP loci. SSR analysis, firstly adopted to classify the clones to their own cultivar, was able to detect ten genetic profiles (one for each cultivar), but not able to distinguish among biotypes of the same cultivar.

As expected, the ampelographic analysis (42 OIV descriptors) was able to well discriminate among cultivars (Fig. 1 and Table S2). In addition, the OIV descriptors resulted different among biotypes of the same cultivar, up to 14 descriptors were variable within a cultivar. Therefore, the Catarratto biotypes resulted clearly distinguished by 14 OIV descriptors (Fig. 1). The density of bunch (OIV 204), the most variable descriptor, was able to distinguish among biotypes belonging to the same other cultivars (Fig. 1). These evidences are consistent with the definition of biotype and that human beings are inclined to select plants for morphological and agronomical differences that finally could influence the qualitative properties of grapes, as occurred for the selection and maintenance of plants with fleshy and large berries, or with white berries (This et al. 2006). However, although the analysis of 42 OIV descriptors was useful to discriminate among biotypes of the same cultivar, these markers are time-consuming and largely influenced by environment (Tessier et al. 1999). The highest morphological variability observed in the cultivar Catarratto was confirmed by genetic analysis that, in contrast, was not able to distinguish sub-clusters including clones belonging each biotype.

Thus, high-throughput SNP genotyping could provide an additional tool to study the genetic diversity and the population structure of the Sicilian grapevine germplasm. The Vitis18kSNP array, developed through NGS technologies, represents a very useful tool to discover genome-wide allelic variation for genetic diversity that could replace the SSR markers for cultivar identification. After removing SNPs having a range of NC from 20 to 100 %, the analysis were carried out on the amplified loci showing a GT score lower than 0.6, as well, providing a good coverage of whole genome (79 %). Since the Vitis18kSNP array contains about 25 % of loci identified from different *Vitis* species (*V. aestivalis*, *V. berlandierii*, *V. labrusca*, *V. cinerea*, *V. lincecumi*, and *Muscadinia rotundifolia*), the percentage of SNP loci (3 %) showing any fragment amplification appeared reasonable, as compared to previous reports (Bekele et al. 2013; De Lorenzis et al. 2015). The percentage of polymorphic loci was high, and the values of heterozygosity (expected, H_e and observed, H_o), very similar among them, lower than those reported for Sicilian collections analyzed by SSR markers (Carimi et al. 2010; De Lorenzis et al. 2014). These results are expected due to the bi-allelic nature of SNPs, although the higher discriminating powerful through the larger number of loci analyzed. Similar results were reported analyzing 700 grapevine cultivars by both 22 SSRs and 384 SNPs (Emanuelli et al. 2013). The MAF mean values were also comparable with those reported by Lijavetzky et al. (2007), who analyzed a collection of about 300 *V. vinifera* accessions (MAF = 0.24), and Emanuelli et al. (2013) (MAF = 0.25). The negative value of F was consistent with high heterozygosity values, meaning an excess of heterozygosity due to a probably prevalence of outcrossing.

As revealed by SSR, SNP analysis confirmed the properly classification of each clone (101) to its own cultivar (Figs. 2 and 4). In addition, SNP polymorphisms among plants of the same cultivar were detected, ranged from 8 (Nerello Cappuccio) to 247 (Catarratto) loci. Unfortunately, these polymorphisms were not able to classify each biotype belonging to the same cultivar in the same cluster, underlining a lack of correlation between genetic and morphological diversity. As example, the three biotypes of Catarratto showed marked differences in morphological (Table S2) and agronomical traits (data not shown) and large variability in their SNP profiles, not able to distinguish among biotype A and B (Catarratto Bianco Comune) and biotype C (Catarratto Bianco Lucido), as previously reported by Crespan et al. (2008), based on SSR analysis. The same authors reported the synthesis of epicuticle waxes covering berry skin as the unique discriminating trait among Catarratto biotypes (Crespan et al. 2008); thus, we concluded that neither the chosen SSR nor SNP loci were able to cover this mutation. In contrast, one or few SNP loci have been reported to discriminate among clones of the same cultivar, such as the SNP in the DXS locus of Muscat and non-Muscat aromatic cultivars (Emanuelli et al. 2010) or the SNP in GAI gene, which determines the number of leaf hairs and reduces plant height and promotes flowering (Boss and Thomas 2002). More recently, the sequencing of Pinot noir clones showed that highly polymorphic *Gypsy*-like elements

are the major causes (about 85 % of the total polymorphic sites) in mutational events occurred in somatic mutations, followed by SNPs (11 %) and indels (4 %) (Carrier et al. 2012). Based on these evidences, the role of the *Gret1* insertion retrotransposon in the promoter region of *VvMybA1* gene to determine the absence of color in the berry skin were demonstrated (Kobayashi et al. 2004; Yakushiji et al. 2006; Vezzulli et al. 2012). Further, the insertion of *Hatvine1-rrm* transposable element in the *VvTFL1A* promoter was reported to cause differences in cluster shape in cultivar Carignan (Fernandez et al. 2010).

As highlighted above, the SNP-based genetic relationship among cultivars identified by cluster analysis (Fig. 2b) supported their distribution underlined also in the PCoA (Fig. 2a), except the Pinot Noir, that in the dendrogram appeared as the most divergent cultivar compared to the others, in agreement with their French origin (Bowers et al. 1999b).

The fastSTRUCTURE analysis inferred three groups based on SNP dataset (Fig. 3), where the largest ones included six Sicilian cultivars. The genetic structure was not able to discriminate between cultivars from Western (Catarratto, Grecanico, Grillo, Inzolia, Perricone, and Zibibbo) to Eastern (Carricante, Frappato, Nerello Cappuccio, and Nero d'Avola) areas of Sicily as already reported (De Lorenzis et al. 2014). Therefore, at $K = 3$, two significant groups of related genotypes were distinguished: Frappato, Perricone, and Zibibbo in the first group (purple) and Grillo, Catarratto, and Grecanico in the second one (blue). The third different genetic structure was assigned to Nero d'Avola, as already observed in the PCoA, the most important and widespread red berry cultivar in Sicily, in agreement with its presumable origin (Calabria).

The admixed genetic structure of Inzolia appeared in accordance with the most known hypothesis about its origin and spread around the Mediterranean Basin, resulting as an important evidence of our analysis. Indeed, molecular evidences already supported the hypothesis that Inzolia, alias Ansonica, was firstly introduced in Sicily by the Greeks in the fourth century B.C. and then spread out in the Island of Giglio (in front of Tuscany) (Labra et al. 1999). Thus, the genetic structure of Inzolia could be the result of human-mediated exchanges between Greece and Magna Graecia throughout history. Greek people domination could have influenced the genetic structure of grapevine varieties by the introduction of foreign varieties utilized for genetic improvement that gave raises the autochthonous cultivars of Sicily (Pastena 2009).

Parentage analysis highlighted significant relationships among cultivars, of which some confirmed previous reports as the cross "Catarratto × Zibibbo" from which Grillo derived (Di Vecchi-Staraz et al. 2007; Cipriani et al. 2010; De Lorenzis et al. 2014). In addition, first-degree relationships between Catarratto and Grecanico (Di Vecchi-Staraz et al. 2007;

Lacombe et al. 2013), as well as the first-degree relationship of Sangiovese with Frappato and Perricone were confirmed (Di Vecchi-Staraz et al. 2007; Gasparro et al. 2013; De Lorenzis et al. 2014). Finally, a first evidence of a second-degree relationship between Catarratto and Nero d'Avola and Inzolia and Nero d'Avola, probably due to their same pedigree, were also found (Fig. 5).

The availability of genome sequence and the high-throughput genotyping platforms have enabled a wide range of applications to clarify the relationships between genotype and phenotype (Yang et al. 2011). Recently, Carrier et al. (2012) presented the first genome-wide analysis of polymorphism among clones of Pinot Noir to identify polymorphisms involved in somatic mutations. Another study, using the Sanger shotgun sequencing and highly efficient sequencing by synthesis (SBS), resolved the complex heterozygous genome, isolating a set of mapped marker loci useful for breeding programs. Cabezas et al. (2011) using a resequencing strategy in selected genotypes developed a set of 48 stable SNP markers with a uniform genome distribution to use for grapevine genotyping.

In summary, single mutation (SNP) and transposon elements can generate somatic variation in grapevine; therefore, the new available high-throughput approaches, such as SNP-array genotyping, RAD-SEQ, and GBS, are very powerful technologies to investigate inter-varietal diversity and population structure of local variety. Nevertheless, in some cases, based on the genome complexity and given the difficulty to identify the different biotypes within a specific cultivar, the integration of different methods might be the best, but more expensive, approach.

Finally, the present study was able to identify, by GAPIT analysis, private SNP profiles related to each cultivar analyzed. In particular, a set of 12 highly polymorphic SNPs, scattered across the genome, can discriminate the main Sicilian cultivars that showed private 12 SNP profiles (Table 3). Specific SNP profiles were able to discriminate all Sicilian cultivars and the reference cultivars. The quality and repeatability of the SNP panel were evaluated by Sanger method.

Conclusion

In this paper, the genetic diversity of ten widespread Sicilian grapevine cultivars was assessed by 42 OIV descriptors, 9 standard SSRs, and the Vitis18kSNP array. The OIV descriptors were utilized for cultivar and biotype morphological characterization. The SNP array was then adopted for genotyping 101 clones from 21 biotypes belonging to the 10 cultivars.

OIV descriptors and SNP datasets were able to distinguish among cultivars, while the recognition among biotypes belonging to the same cultivar appeared more complex. In the next future, large efforts should be devoted to

the analysis of location and function of each polymorphic SNP among biotypes of the same cultivar. Particular attention will be payed to Catarratto that revealed the larger intra-varietal genetic diversity. Although both classes of markers were informative, ampelographic analysis is time-consuming and largely influenced by environment, thus can be replaced by SNP-array. Up to date, for lab automation and cost-effectiveness, the SNP array will represent a very useful tool to investigate the genetic diversity. The development of SNP databases for grapevine cultivars could help the overcoming of SSR also for the true-to-typeness cultivar assignment. Cluster and parentage analyses confirmed a high number of genetic relationships among Sicilian cultivars, based on the Vitis18kSNP array. These results demonstrated that the selection practise was made over the years in Sicily, leading to increase the genetic diversity of grapevine germplasm, to date considered the biggest and oldest winegrowing Italian region.

Finally, the panel of 12 SNPs scattered across the genome can be proposed for a fast and low cost genotyping system to recognize and safeguard the Sicilian grapevine cultivars. This study could represent a starting point to implement and extend the same system to other national and international grapevine cultivar collections.

Authors' contribution   All the authors conceived and designed the experiments; FM and GDL carried out the experiments and performed the statistical analysis; FM, GDL, AL, MGB, and FS commented the results and drafted the manuscript. All the authors critically read and approved the final manuscript.

Data Archiving Statement   We followed standard Tree Genetics and Genomes policy. All SNP data used in this work will be archived on the Italian Vitis Database (http://www.vitisdb.it/) and the International Grape Genome Program (IGGP) (http://www.vitaceae.org). All the accessions are detailed in supplementary file, Table S1.

Compliance with ethical standards

References

Barkley NA, Roose ML, Krueger RR, Federic CT (2006) Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRs). Theor Appl Genet 112:1519–1531

Bekele WA, Wieckhorst S, Friedt W, Snowdon RJ (2013) High-throughput genomics in sorghum: from whole genome resequencing to a SNP screening array. Plant Biotechnol J 11:1112–1125

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc Series B (Methodological) 57(1):289–300

Boss PK, Thomas MR (2002) Association of dwarfism and floral induction with a grape 'green revolution' mutation. Nature 416:847–850

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

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

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

Cabezas JA, Ibáñez J, Lijavetzky D, Vélez D, Bravo G, Rodríguez V, Carreño I, Jermakow AM, Carreño J, Ruiz-García L, Thomas MR, Martinez-Zapater JM (2011) A 48 SNP set for grapevine cultivar identification. BMC Plant Biol 11:153

Campostrini F, De Micheli L, Bogoni M, Scienza A (1995) Study of genetic variability of Sangiovese ecotypes as a tool for new strategies in clonal selection. Proceedings of the International Symposium on Clonal Selection: June 20 & 21, Oregon Convention Centre, Portland, Oregon, Usa, pag. 105–110

Carimi F, Mercati F, Abbate L, Sunseri F (2010) Microsatellite analyses for evaluation of genetic diversity among Sicilian grapevine cultivars. Genet Resour Crop Evol 57:703–719

Carimi F, Mercati F, De Michele R, Fiore MC, Riccardi P, Sunseri F (2011) Intra-varietal genetic diversity of the grapevine (*Vitis vinifera* L) cultivar 'Nero d'Avola' as revealed by microsatellite markers. Genet Resour Crop Evol 58:967–975

Carrier G, Le Cunff L, Dereeper A, Legrand D, Sabot F, Bouchez O, Audeguin L, Boursiquot JM, This P (2012) Transposable elements are a major cause of somatic polymorphism in *Vitis vinifera* L. PLoS ONE 7, e32973

Castro I, D'Onofrio C, Martin JP, Ortiz JM, De Lorenzis G, Ferreira V, Pinto-Carnide O (2012) Effectiveness of AFLPs and retrotransposon-based markers for the identification of Portuguese grapevine cultivars and clones. Mol Biotechnol 52:26–39

Cervera MT, Cabezas JA, Sancha JC, Martinez de Toda F, Martinez-Zapater JP (1998) Application of AFLPs to the characterisation of grapevine *Vitis vinifera* L. genetic resources. A case study with accessions from Rioja (Spain). Theor Appl Genet 97:51–59

Chagné D, Crowhurst RN, Troggio M, Davey MW, Gilmore B, Lawley C, Vanderzande S, Hellens RP, Kumar S, Cestaro A, Velasco R, Main D, Rees JD, Iezzoni A, Mockler T, Wilhelm L, Van de Weg E, Gardiner SE, Bassil N, Peace C (2012) Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. PLoS ONE 7, e31745.24

Cipriani G, Marrazzo MT, Marconi R, Cimato A, Testolin R (2002) Microsatellite markers isolated in olive (*Olea europaea* L.) are suitable for individual fingerprinting and reveal polymorphism within ancient cultivars. Theor Appl Genet 104:223–228

Cipriani G, Spadotto A, Jurman I, Di Gaspero G, Crespan M, Meneghetti S, Frare E, Vignani R, Cresti M, Morgante M, Pezzotti M, Pe E,

Policriti A, Testolin R (2010) The SSR-based molecular profile of 1005 grapevine (*Vitis vinifera* L.) accessions uncovers new synonymy and parentages, and reveals a large admixture amongst varieties of different geographic origin. Theor Appl Genet 121:1569–1585

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

Cretazzo E, Meneghetti S, De Andrés MT, Frare E, Gaforio L, Cifre J (2010) Clone differentiation and varietal identification by means of SSR, AFLP, SAMPL and M-AFLP in order to assist the clonal selection of grapevine. The case of study of Manto Negro, Callet and Moll, autochthonous cultivar of Majorca. Ann Appl Biol 157: 213–227

D'Amato F (1997) Role of somatic mutations in the evolution of higher plants. Caryologia 50:1–15

D'Onofrio C, De Lorenzis G, Giordani T, Natali L, Cavallini A, Scalabrelli G (2010) Retrotransposon-based molecular markers for grapevine species and cultivars identification. Tree Genet Genomes 6:451–466

De Lorenzis G, Imazio S, Biagini B, Failla O, Scienza A (2013) Pedigree Reconstruction of the Italian Grapevine Aglianico (*Vitis vinifera* L.) from Campania. Mol Biotechnol 54:634–642

De Lorenzis G, Las Casas G, Brancadoro L, Scienza A (2014) Genotyping of Sicilian grapevine germplasm resources (*V. vinifera* L.) and their relationships with Sangiovese. Sci Hortic 169:189–198

De Lorenzis G, Chipashvili R, Failla O, Maghradze D (2015) Study of genetic variability in *Vitis vinifera* L. germplasm by high-throughput Vitis18kSNP array: the case of Georgian genetic resources. BMC Plant Biol 15:154

Dereeper A, Nicolas S, Le Cunff L, Bacilieri R, Doligez A, Peros JP, Ruiz M, This P (2011) SNiPlay: a web-based tool for detection, management and analysis of SNPs. Application to grapevine diversity projects. BMC Bioinformatics 12:134

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

Dirlewanger E, Cosson P, Tavaud M, Aranzana M, Poizat C, Zanetto A, Arus P, Laigret F (2002) Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). Theor Appl Genet 105:127–138

Emanuelli F, Battilana J, Costantini L, Le Cunff L, Boursiquot JM, This P, Grando MS (2010) A candidate gene association study on muscat flavor in grapevine (*Vitis vinifera* L.). BMC Plant Biol 10:241

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

Fernandez L, Torregrosa L, Segura V, Bouquet A, Martinez-Zapater JM (2010) Transposon-induced gene activation as a mechanism generating cluster shape somatic variation in grapevine. Plant J 61:545–557

Forni G (2012) The origin of "Old World" viticulture. In: Maghradze D, Rustioni L, Scienza A, Turok J, Failla O (eds) Caucasus and Northern Black Sea Region, 1st edn. JKI - Julius Kühn-Institut: Vitis Special Issue, Geilweilerhof, pp 27–38

Ganal MW, Polley A, Graner EM, Plieske J, Wieseke R, Luerssen H, Durstewitz G (2012) Large SNP arrays for genotyping in crop plants. J Biosci 37:821–828

Gardner KM, Brown P, Cooke TF, Cann S, Costa F, Bustamante C, Velasco R, Troggio M, Myles S (2014) Fast and cost-effective genetic mapping in apple using next-generation sequencing. G3: Genes|Genomes|Genetics 4:1681–1687

Gasparro M, Caputo AR, Bergamini C, Crupi P, Cardone MF, Perniola R, Antonacci D (2013) Sangiovese and its offspring in Southern Italy. Mol Biotechnol 54:581–589

González Techera A, Jubany S, Ponce De León I, Boido E, Dellacassa E, Carrau FM, Hinrichsen P, Gaggero G (2004) Molecular diversity within clones of cv. Tannat (Vitis vinifera). Vitis 43:179–185

Jaillon O, Aury JM, 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, Bruyère C, Billault A, Segurens B, Gouyvenoux B, Ugarte E, Cattonaro F, Anthouard F, 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, Pè ME, 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 449:463–467

Jiang Z, Wang H, Michal JJ, Zhou X, Liu B, Woods LC, Fuchs RA (2016) Genome wide sampling sequencing for SNP genotyping: methods, challenges and future development. Int J Biol Sci 12: 100–108

Kobayashi S, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-induced mutations in grape skin color. Science 304:982–982

Kurokawa Y, Noda T, Yamagata Y, Angeles-Shim R, Sunohara H, Uehara K, Furuta T, Nagai K, Jena KK, Yasui H, Yoshimura A, Ashikari M, Doi K (2016) Construction of a versatile SNP array for pyramiding useful genes of rice. Plant Sci 242:131–139

Labra M, Failla O, Fossati T, Castiglione S, Scienza A, Sala F (1999) Phylogenetic analysis of grapevine cv. Ansonica growing on the island of Giglio, Italy, by AFLP and SSR markers. Vitis 38:161–166

Lacombe T, Boursiquot JM, Laucou V, Di Vecchi-Staraz M, Péros JP, This P (2013) Large-scale parentage analysis in an extended set of grapevine cultivars (*Vitis vinifera* L.). Theor Appl Genet 126:401–414

Laidò G, Mangini G, Taranto F, Gadaleta A, Blanco A, Cattivelli L, Marone D, Mastrangelo AM, Papa R, De Vita P (2013) Genetic diversity and population structure of tetraploid wheats (*Triticum turgidum* L.) estimated by SSR, DArT and pedigree data. PLoS ONE 8(6), e67280

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

Le Paslier M-C, Choisne N, Bacilieri R, Bounon R, Boursiquot J-M, Bras M, Brunel D, Di Gaspero G, Hausmann L, Lacombe T, Laucou V, Launay A, Martinez-Zapater JM, Morgante M, Raj PM, Ponnaiah M, Quesneville H, Scalabrin S, Torres-Perez R, Adam-Blondon A-F: The GrapeReSeq 18 k Vitis genotyping chip. In: 9th International symposium grapevine physiology and biotechnology: International Society for Horticultural Science; 21–26 April 2013; La Serena, p. 123

Lijavetzky D, Cabezas JA, Ibáñez A, Rodríguez V, Martínez-Zapater JM (2007) High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. BMC Genomics 8:424

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: Genome Association and Prediction Integrated Tool. Bioinformatics 28:2397–2399

Maitti C, Andreani L, Geuna F, Brancadoro L, Scienza A (2009) Genetic characterization of *Vitis vinifera* accessions cultivated in Sicily (Italy). Acta Hortic 827:177–182

Maul E, Sudharma KN, Kecke S, Marx G, Müller C, Audeguin L, Boselli M, Boursiquot JM, Bucchetti B, Cabello F, Carraro R, Crespan M, de Andrés MT, Eiras Dias J, Ekhvaia J, Gaforio L, Gardiman S, Grando S, Argyropoulos D, Jandurova O, Kiss E, Kontic J, Kozma P, Lacombe T, Laucou V, Legrand D, Maghradze D, Marinoni D, Maletic E, Moreira F, Muñoz-Organero F, Nakhutsrishvili G, Pejic I, Peterlunger E, Pitsoli D, Pespisilova D, Preiner D, Raimondi S, Regner F, Savin G, Savvides S, Schneider A, Sereno C, Simon C, Staraz M, Zulini L, Bacilieri R, This P (2012) The European Vitis Database (www.eu-vitis.de): a technical innovation through an online uploading and interactive modification system. Vitis 51:79–85

Melo ATO, Bartaula R, Hale I (2016) GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. BMC Bioinformatics 17:29

Mercati F, Riccardi P, Harkess A, Sala T, Abenavoli MR, Leebens-Mack J, Falavigna A, Sunseri F (2015) Single nucleotide polymorphism-based parentage analysis and population structure in garden asparagus, a worldwide genetic stock classification. Mol Breed 35:59

Moreno S, Gogorcena Y, Ortiz JM (1995) The use of RAPD markers for identification of cultivated grapevine (Vitis vinifera L.). Sci Hortic 62:237–243

Moreno S, Martin J, Ortiz J (1998) Inter simple sequence repeats PCR for characterization of closely related grapevine germplasm. Euphytica 101:117–125

Myles S, Chia JM, Hurwitz B, Simon C, Zhong GY, Buckler E, Ware D (2010) Rapid genomic characterization of the genus Vitis. PLoS ONE 5, e8219

Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F, Aradhya MK, Prins B, Reynolds A, Chia J-M, Ware D, Bustamante CD, Buckler ES (2011) Genetic structure and domestication history of the grape. Proc Natl Acad Sci U S A 108:3530–3535

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

Okitsu CY, Van Den Berg DJ, Lieber MR, Hsieh CL (2013) Reproducibility and reliability of SNP analysis using human cellular DNA at or near nanogram levels. BMC Res Notes 6:515

Pastena B (2009) La civiltà della vite in Sicilia. La vitivinicoltura siciliana nel tempo. Edizioni Leopardi, Palermo

Peace C, Bassil N, Main D, Ficklin S, Rosyara UR, Stegmeir T, Sebolt A, Gilmore B, Lawley C, Mockler TC, Bryant DW, Wilhelm L, Iezzoni A (2012) Development and evaluation of a genome-wide 6K SNP array for diploid sweet cherry and tetraploid sour cherry. PLoS ONE 7, e48305

Pelsy F, Hocquigny S, Moncada X, Barbeau G, Forget D, Hinrichsen P, Merdinoglu D (2010) An extensive study of the genetic diversity within seven French wine grape variety collections. Theor Appl Genet 120:1219–1231

Pindo M, Vezzulli S, Coppola G, Cartwright DA, Zharkikh A, Velasco R, Troggio M (2008) SNP high-throughput screening in grapevine using the SNPlex^{TM} genotyping system. BMC Plant Biol 8:12

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575

Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics 197:573–589

Reif JC, Warburton ML, Xia XC, Hoisington DA, Crossa J, Taba S, Muminovic J, Bohn M, Frisch M, Melchinger AE (2006) Grouping of accessions of Mexican races of maize revisited with SSR markers. Theor Appl Genet 113:177–185

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138

Schmid KJ, Sorensen TR, Stracke R, Torjek O, Altmann T, Mitchell-Olds T, Weisshaar B (2003) Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in Arabidopsis thaliana. Genome Res 13:1250–1257

Sefc KM, Regner F, Turetschek E, Glossl J, Steinkellner H (1999) Identification of microsatellite sequences in Vitis riparia and their applicability for genotyping of different Vitis species. Genome 42:367–373

Sim SC, van Deynze A, Stoffel K, Douches DS, Zarka D, Ganal MW, Chetelat RT, Hutton SF, Scott JW, Gardner RG, Panthee DR, Mutschler M, Myers JR, Francis DM (2012) High-density SNP genotyping of tomato (Solanum lycopersicum L.) reveals patterns of genetic variation due to breeding. PLoS ONE 7, e45520

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Tayeh N, Aluome C, Falque M, Jacquin F, Klein A, Chauveau A, Bérard A, Houtin H, Rond C, Kreplak J, Boucherot K, Martin C, Baranger A, Pilet-Nayel ML, Warkentin TD, Brunel D, Marget P, Le Paslier MC, Aubert G, Burstin J (2015) Development of two major resources for pea genomics: the GenoPea 13.2K SNP Array and a high-density, high-resolution consensus genetic map. Plant J 84:1257–1273

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

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

Thomas MR, Scott NS (1993) Microsatellite repeats in grapevine reveal DNA polymorphism when analysed as sequence-tagged sites (STSs). Theor Appl Genet 86:985–990

Thomson MJ, Septiningsih EM, Suwardjo F, Santoso TJ, Silitonga TS, McCouch SR (2007) Genetic diversity analysis of traditional and improved Indonesian rice (Oryza sativa L.) germplasm using microsatellite markers. Theor Appl Genet 114:559–568

Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, FitzGerald LM, Vezzulli S, Reid J, Malacarne G, Iliev D, Coppola G, Wardell B, Micheletti D, Macalma T, Facci M, Mitchell JT, Perazzolli M, Eldredge G, Gatto P, Oyzerski R, Moretto M, Gutin N, Stefanini M, Chen Y, Segala C, Davenport C, Demattè L, Mraz A, Battilana J, Stormo K, Costa F, Tao Q, Si-Ammour A, Harkins T, Lackey A, Perbost C, Taillon B, Stella A, Solovyev V, Fawcett JA, Sterck L, Vandepoele K, Grando SM, Toppo S, Moser C, Lanchbury J, Bogden R, Skolnick M, Sgaramella V, Bhatnagar SK, Fontana P, Gutin A, Van de Peer Y, Salamini F, Viola R (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS ONE 2, e1326

Verde I, Bassil N, Scalabrin S, Gilmore B, Lawley CT, Gasic K, Micheletti D, Rosyara UR, Cattonaro F, Vendramin E, Main D, Aramini V, Blas AL, Mockler TC, Bryant DW, Wilhelm L, Troggio M, Sosinski B, Aranzana MJ, Arús P, Iezzoni A, Morgante M, Peace C (2012) Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. PLoS ONE 7, e35668

Vezzulli S, Leonardelli L, Malossini U, Stefanini M, Velasco R, Moser C (2012) Pinot blanc and Pinot gris arose as independent somatic mutations of Pinot noir. J Exp Bot 63:6359–6369

Vivier M, Pretorius IS (2002) Genetically tailored grapevines for the wine industry. Trends Biotechnol 20:472–478

Winfield MO, Allen AM, Burridge AJ, Barker GLA, Benbow HR, Wilkinson PA, Coghill J, Waterfall C, Davassi A, Scopes G, Pirani A, Webster T, Brew F, Bloor C, King J, West C, Griffiths S, King I, Bentley AR, Edwards KJ (2015) High-density SNP genotyping

array for hexaploid wheat and its secondary and tertiary gene pool. Plant Biotechnol J. doi:10.1111/pbi.12485

Xu S, Gupta S, Jin L (2010) PEAS V10: a package for elementary analysis of SNP data. Mol Ecol Res 10:1085–1088

Yakushiji H, Kobayashi S, Goto-Yamamoto N, Jeong ST, Sueta T, Mitani N, Azuma A (2006) A skin color mutation of grapevine, from black-skinned 'Pinot Noir' to white-skinned 'Pinot Blanc' is caused by the deletion of the functional *VvMybA1* allele. Biosci Biotechnol Biochem 70:1506–1508

Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, de Andrade M, Feenstra B, Feingold E, Hayes MG, Hill WG, Landi MT, Alonso A, Lettre G, Lin P, Ling H, Lowe W, Mathias RA, Melbye M, Pugh E, Cornelis MC, Weir BS, Goddard ME, Visscher PM (2011) Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet 43:519–525

Yu J, Pressoir G, Briggs W, Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Yu H, Xie W, Li J, Zhou F, Zhang Q (2014) A whole-genome SNP array (RICE6K) for genomic breeding in rice. Plant Biotechnol J 12:28–37

Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–360

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326–3328

Zohary D, Horf M, Weiss E (2000) Domestication of plants in the Old World, 3rd edn. Oxford University Press, Oxford