Single nucleotide polymorphism profiles reveal an admixture genetic structure of grapevine germplasm from Calabria, Italy, uncovering its key role for the diversification of cultivars in the Mediterranean Basin

F. SUNSERI[1], A. LUPINI[1], A. MAUCERI[1], G. DE LORENZIS[2], F. ARANITI[1], L. BRANCADORO[2], A. DATTOLA[1], G. GULLO[1], R. ZAPPIA[1] and F. MERCATI[3]

[1] Dipartimento AGRARIA, Università Mediterranea di Reggio Calabria, I-89121, Reggio Calabria, Italy; [2] Dipartimento di Scienze Agrarie ed Ambientali, Università degli Studi di Milano, I-20133, Milano, Italy; [3] Istituto di Bioscienze e Biorisorse (IBBR), Consiglio Nazionale delle Ricerche (CNR), I-90120, Palermo, Italy.

Corresponding authors: Professor Francesco Sunseri, emailfrancesco.sunseri@unirc.it and Dr Francesco Mercati, email francesco.mercati@ibbr.cnr.it

Short title: Genetic diversity of grapevines from Calabria, Italy

Abstract

Background and Aims: Grapevine (*Vitis vinifera* L.) is one of the earliest domesticated crops, dating back 6000 years ago in the Near East before spreading into Europe. Despite the abundance of historical, archaeological, and genetic records, until now the following domestication events are not fully understood. Here, a genetic characterisation of grape germplasm from Calabria (Southern Italy), a crucial area of the Mediterranean Basin, aims to validate this area as secondary centre of crop domestication.

Methods and Results: True-to-type cultivar classification of 72 accessions was carried out by using microsatellite loci (SSR) along with the main OIV ampelographic descriptors. A high level of genetic diversity (H_e= 0.83) among native cultivars from Calabria was highlighted. A decay of genetic diversity moving from Southern Italy to North-Western Mediterranean regions was observed, probably due to repeated founder effects during the grapevine expansion from Mediterranean Basin to Europe. Single nucleotide polymorphism (SNP) analysis showed an admixture genetic structure at K= 7, clarifying a complex network of pedigree relationships generated by crosses among cultivars. Parentage analysis underlined a high proportion of parent-offspring relationships (76%) in Mantonico Bianco and Pecorello, hypothesising their key role in many native cultivars pedigree from Southern Mediterranean area.

Conclusion: Overall, our results appear to indicate a pivotal role of cultivars from Calabria in the grape genetic diversity of Southern Italy. Further, genetic analysis on grape wild accessions from Calabria should be useful to discuss a secondary centre of grape domestication.

Significance of the Study: A large grape collection from Calabria was for the first time characterised through ampelographic and genetic analyses.

Keywords: *ampelography, domestication centre, genetic diversity, native cultivars, parentage analysis, SSR simple sequence repeat, single nucleotide polymorphism,* Vitis vinifera *L.*

Introduction

The grapevine (*Vitis vinifera* ssp. *vinifera*) is one of the oldest and most economically important crops in the world (Vivier and Pretorius 2002). It is a clonally propagated and highly heterozygous crop distributed from Central Asia to the Mediterranean Basin (Zohary et al. 2000). Based on archaeological record, cultivated grapevine is thought to have been domesticated from the wild progenitor *Vitis vinifera* ssp. *sylvestris* in the Near East regions between the Black Sea and Iran in the second half of the 4th millennium BC (Zohary and Hopf 1993, McGovern 2003). From this area, cultivated forms were spread by humans to the Mediterranean regions, in Greece, Crete and Cyprus. The first testimonies of grapevine cultivation in Southern Italy were dated in the 2nd millennium BC (Di Vora and Castelletti 1995). then spread to Spain and Maghreb, due to the Phoenician influence during the first part of the last millennium BC (Rivera Nunez and Walker 1989, Buxò 2008), and in France by the Greek Phocaeans with the foundation of Marseille [600 B.C.; Brun and Laubenheimer (2001)].

Among Southern Italian regions, Calabria is one of the most important and ancient for grape cultivation, with a long tradition of viticulture practices. The myth refers that Calabria's antique denomination *Enotria – Terra del vino* (Land of wine) from Greek *oinos* (wine) was due to a leader of Arcadians, founder of an ancient colony (Enotri) in Southern Italy around 1600 BC (Teti 2008). The most antique literary testimony of wine production in that area was from Strabone (VI, 1, 14) and Plinio il Vecchio (*Naturalis Historia* XIV, 8, 89) in the first century BC in which a wine from *Thurii* (today area of Castrovillari, Northern Calabria) was cited (Teti 2008). The introduction of phylloxera in Europe at the end of 19th century caused major destruction even in the vineyards of Calabria. The replanting during the first decades of the 20th century with American vine rootstock, resistant to the pest, changed the Calabrian wine scenario forever, either for the cultivars or the reduced number of hectares cultivated. Fortunately, many of the native varieties, always cultivated in Calabria, were recovered from relict vineyards to be sprayed in the regional grapevine districts.

Domestication processes and outbreeding determined a vast genetic variation among grapevine cultivars around the word. Grape genome accumulated deletions, insertions, inversions, single nucleotide polymorphisms (SNP) (This et al. 2006, Jaillon et al. 2007, Velasco et al. 2007) that together with natural or artificial crossing were the drivers of grapevine evolution since its domestication (This et al. 2006, Forni 2012). Therefore, more than 10 000 cultivars are believed to exist around the word (Alleweldt and Possingham 1988).

The grapevine genetic diversity can be assessed by morphological traits (Alleweldt and Possingham 1988) and ampelography, useful to identify many grape cultivars (Galet 1979, Galet

1991, IPGRI UPOV Organisation Internationale de la Vigne et du Vin 1997). Unfortunately, these traits can be influenced by the environment (Levadoux 1956, Tessier et al. 1999). To distinguish better among grapevine cultivars, different DNA-based markers, among which amplified (AFLP) (Cervera et al. 1998), ISSR (Moreno et al. 1998) and simple sequence repeat (SSR) (Bowers et al. 1996, Sefc et al. 1999), were used. Simple sequence repeat or microsatellites consist of tandemly repeated sequence motifs with a high variation in repeat number among individuals (Morgante and Olivieri 1993). Simple sequence repeat are co-dominant, multi-allelic and highly reproducible by standardisation, playing a key role to evaluate genetic diversity in crops, such as wheat (Laidò et al. 2013) and *Citrus* (Barkley et al. 2006). Also grape genotyping is mainly based on SSR, for cultivar identification, finding relationships among cultivars, synonyms, homonyms (Cipriani et al. 2010, Carimi et al. 2010, Laucou et al. 2011, Emanuelli et al. 2013), parentage analysis (Lacombe et al. 2013) and population structure (Biagini et al. 2014). Simple sequence repeat showed their limit, being not always able to discriminate among clones/biotypes of the same cultivar (González-Techera et al. 2004, Pelsy et al. 2010).

The development of high-throughput next generation sequencing (NGS) technologies has made possible to sequence entire genomes more efficiently, obtaining large-scale SNP (single nucleotide polymorphism) isolation and the advance of efficient SNP genotyping platforms (Schmid et al. 2003, Lijavetzky et al. 2007, Pindo et al. 2008, Verde et al. 2012, Yu et al. 2014, Melo et al. 2016) useful for detect genetic variations (Myles et al. 2011, Sim et al. 2012, Winfield et al. 2015, Kurokawa et al. 2016, Mercati et al. 2016). Large-scale grape SNP discovery has been reported (Lijavetzky et al. 2007, Pindo et al. 2008) leading to the identification of thousands of SNP validated in a 9K genotyping array (Myles et al. 2010, 2011). More recently, the GrapeReSeq Consortium developed a high-throughput 18K SNP chip (Le Paslier et al. 2013) and informative SNPs set for cultivar identification and clonal variation studies were also identified (Cabezas et al. 2011, Carrier et al. 2012, Mercati et al. 2016).

In the present paper, a grapevine collection from Calabria was characterised for the first time by the main OIV ampelographic descriptors and both nuclear SSR (selecting nine among the most informative) and the high-throughput Vitis18kSNP array (Le Paslier et al. 2013). The main goals were finalised to: (i) assess the genetic structure of germplasm from Calabria; (ii) discriminate among the existing synonymies and homonymies described from many vine-growing regional districts; (iii) identify genetic relationships among cultivars by parentage analysis; and (iv) evaluate the level of genetic diversity compared to grapevine germplasm of the North-Western Mediterranean regions (i.e. North-Centre of Italy, Spain and France). Our results well distinguished the main cultivars from

Calabria, providing additional information about their genetic relationships by parentage analysis, underlining a central role of two cultivars (Mantonico Bianco and Pecorello) for the development of many other cultivars from that area. Further, the analysis of amount and distribution of genetic diversity supports the hypothesis of a chief involvement of cultivars from Calabria and Sicily to the origin of current grapevine germplasm, suggesting these regions as candidate for a secondary centre of domestication in the Western Mediterranean region (Grassi et al. 2003, Arroyo-Garcia et al. 2006, De Mattia et al. 2008).

Material and methods

Plant material and DNA extraction

Grape genotypes were collected from different areas of Calabria by the Department of Agraria at the University of Reggio Calabria and the Department of Scienze Agrarie e Ambientali at the University of Milan, some accessions derived from the repository available at the experimental station of Agenzia Regionale Servizi Sviluppo Agricolo (ARSSA) of Calabria at S. Marco Argentano (CS), Italy. Seventy-two grapevine accessions belonging to the cultivars from Calabria were taken into account for the analyses and maintained at the experimental station of the Department of Agraria – Reggio Calabria (Table S1). Pinot Noir and Sangiovese were included in the analysis as international and national reference cultivars, respectively. Genomic DNA was extracted from 100 mg of young leaves tissue (1–2 cm of diameter) or woody tissue, using the QiagenDNeasy Plant Mini Kit (Qiagen, Hilden, Germany). The DNA quality (260/230 and 260/280 ratios) and concentration was checked by NanoDrop Spectrophotometer (Thermo Scientific, Waltham, MA, USA).

Ampelographic analysis

The most cultivated accessions in the local farms were chosen for the ampelographic analysis; three cloned plants from each of 21 out of 72 accessions were utilised. Different accessions belong to Greco Nero = Magliocco Dolce cultivar were included in the analysis because of the high number of synonyms found for this cultivar. Thirty-six ampelographic traits, 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 2013 and 2014. The observations (from six to ten depending on the descriptor) were carried out in different times during the vegetative seasons, as recommended 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).

Finally, using the package R/ggplot2 (https://cran.r-project.org/web/packages/ggplot2/index.html, R package version 1.1), a heatmap describing the variation of OIV descriptors was set up. Each descriptor was recorded on a 1–9 scale, and the different colours and gradients were associated to the scale and combination for each category.

SSR and SNP genotyping

All the 72 accessions were analysed by using nine SSR (VrZag62, VrZag79, VVMD5, VVMD7, VVMD25, VVMD27, VVMD28, VVMD32, VVS2) (Thomas and Scott 1993, Bowers et al. 1996, 1999, Sefc et al. 1999), suggested as a standard set for grapevine genotyping in the frame of the GrapeGen06 European project. Amplification reactions were performed in 20 µL final volume reaction mixture as described in Mercati et al. (2013). The amplification fragments were detected on ABI PRISM 3500 Genetic Analyser (Applied Biosystems by Life Technologies, Foster City, CA, USA), and the alleles were sized by GENEMAPPER 4.0 (Applied Biosystems by Life Technologies). Pinot Noir and Sangiovese were included as references for standardisation. Simple sequence repeat profiles were compared to the Italian and European Vitis databases.

SNP genotype data for the panel (72 grapevine accessions from Calabria and the reference cultivars) were generated from the custom Illumina Vitis18KSNP array (Illumina, San Diego, CA, California), which assays 18 071 SNPs (Le Paslier et al. 2013). The DNAs were delivered to TraitGenetics (Gatersleben, Salzlandkreis, Germany) for genotyping. Two hundred nanograms of genomic DNA were used as template for each reaction, following the manufacturer's instructions (Illumina).

Analysis of SNP-data

Row data were visualised and analysed by using GenomeStudio V2011.1 software (Illumina). The dataset was filtered and standardised removing (i) samples with low SNP call quality (*p50GC*<0.54); (ii) SNPs with a GenTrain score lower than 0.6; (iii) monomorphic SNPs; and (iv) SNPs with several non-calls (NCs) higher than 20%. SNPs with MAF > 0.05. and missing rate < 0.20 were used for all developed analyses.

PLINK (http://pngu.mgh.harvard.edu/purcell/plink/; Purcell et al. 2007), the Bioconductor packages snpStats (snpStats: SnpMatrix and XSnpMatrix classes and methods. R package version 1.12.0) and HierFstat (Goudet 2005) were used to compute the main genetic parameters. To investigate the genetic relationships among cultivars studied both cluster analysis and kinship estimation were carried out. A phylogenetic tree was designed by both unweighted pair group method with arithmetic mean (UPGMA) and neighbour joining (NJ) method by using R/adegenet 1.3

(Jombart and Ahmed 2011) and the appropriate method for our dataset has been chosen. The bootstrap analysis was performed based on 100 resampling. The degree of kinship among genotypes was estimated from molecular data in R/GAPIT (Lipka et al. 2012) using VanRaden's method (VanRaden 2008). FastStructure (Raj et al. 2014) was used in admixture model to highlight the number of putative genetic pools (*K*) present in the dataset. The analysis was performed and visualised as described in Mercati et al. (2016), using the input files (.bed, .bim, .fam) generated by PLINK. The package R/SNPRelate (Zheng et al. 2012) was used to evaluate the structuration of the Calabrian cultivars compared to the main Sicilian cultivars previously analysed (Mercati et al. 2016) by computing a principal component analysis (PCA) extracting the first 100 PC. The Wrigth's fixation index (F_{st}) (Wright 1965) was evaluated for each genetic background studied (both Calabria and Sicily) using R/HierFstat. The Excel Microsatellite Toolkit (Park 2001) was used to determine the gene diversity and the mean of polymorphic information content (PIC; Botstein et al. 1980) among cultivars belonging to Italian regions (http://catalogoviti.politicheagricole.it) and other North-Western Mediterranean countries (Lacombe et al. 2013).

Since the higher values of inbreeding coefficient (*F*) would correlate with an increased length of runs of homozygosity (ROH), defined as homozygous haplotypes of various lengths (Magi et al. 2014), using the selected SNPs distributed genome-wide, we also evaluated levels of ROHs. Only SNPs in a 2-kb window were considered in order to identify segments representing homozygosity by descent (i.e. autozygosity) rather than by chance. PLINK was used to identify segments of ROH in a window containing 50 SNPs, among which the number of maximum missing SNPs and number of maximum heterozygous sites was set to 2 and 1, respectively. Cultivars with overlapping ROH, and those ROH that were an allelic match were also identified. The overlapping regions were isolated by using the sliding window approach as mentioned above. Physical distances were converted into genetic distances based on the available *Vitis vinifera* genome (Jaillon et al. 2007, Velasco et al. 2007).

Parentage analysis

Identity-By-Descent (IBD) index for all pairwise comparisons among the 74 *V. vinifera* accessions was carried out using PLINK. MAF and r^2 of LD were set on 0.01 and 0.05 values, respectively. Probability of sharing 0 IBD allele identical-by-descent (Z0), 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 x P (IBD = 1)] were evaluated. In parent-offspring relationships, Z0 and Z2 are expected to be 0 and Z1 to be 1; while in 2nd 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 and pairs of accessions showing IBD >95% were considered genetically identical (clones). Most Likely Relationship (MLRs) were evaluated and visualised using R/circlize (Gu et al. 2014).

Results

Ampelographic analysis

Thirty-six out of 48 OIV descriptors suggested by the European GrapeGen06 project (Maul et al. 2012) were scored twice during spring-summer seasons 2013 and 2014 on the accessions belonging the main grapevine cultivars from Calabria. A detailed report of the 36 OIV descriptors together with the main traits discriminating among cultivars or accessions from the same cultivar is included in Table S2. A heatmap representing the expression level of each OIV descriptor in each cultivar was also provided (Figure 1). The descriptors OIV 4, 51, 53, 68, 72, 74, 75, 76, 79, 84, 204, 206, 228, 225, and 235 showed differences among cultivars that can be clearly distinguished (Figure 1, Table S2). The descriptors showing the greatest difference among the 14 cultivars from Calabria were OIV 4 (young shoot: density of prostrate hairs on the shoot tip), with values ranging from 1 (short) to 9 (very long), OIV 204 (density of bunch), with values ranging from 1 (very loose) to 7 (dense) and OIV 84 (mature leaf: density of prostrate hairs between main veins on lower side of blade), with values ranging from 1 (very loose) to 9 (very dense) overall the samples.

SSR-based true-to-type cultivar classification

Nine SSR markers selected from the European GrapeGen06 project were utilised to establish the genetic profile of each accession belonging to their own cultivar. Twenty-one SSR profiles were obtained, one for each cultivar collected in Calabria and maintained for ampelographic analysis. In addition, other eight SSR unique profiles (data not shown) were found together with the reference cultivars, Pinot Noir and Sangiovese (Table S3). The identity of each genotype (true-to-type) was assessed by SSR profiles comparison with public databases, such as the Italian Register of Grapevine Catalogue (Registro Nazionale Varietà Vite – RNVV) (http://catalogoviti.politicheagricole.it), the Italian (http://www.vitisdb.it) and the European (http://www.vivc.de) Vitis Databases. The SSR profiles were also compared with CREA-VIT (Ente CREA Conegliano Veneto – Italy) database for the classification of genotypes into true type cultivar (Table S3). Accessions holding the same SSR profile are frequently collected with different names (synonymies), vice versa few accessions named as already known cultivars held SSR profile not corresponding to the cultivar (homonymies) (Table

S4). True-to-type cultivars were used as anchors/links in the SNP genotyping.

SNP analysis and genetic diversity

The high-throughput Vitis18kSNP array was used to investigate the genetic relationships among 72 grapevine cultivars representative of Calabrian germplasm, using Pinot Noir and Sangiovese as references. After SNP-dataset filtering, 411 loci (2%) did not amplify among all genotypes and about 10 295 loci (57%) showed GT score higher than 0.6 (Table 1). The final dataset resulted in 10 041 out of 18 071 loci after removing the SNPs with a number of NC (not-call) higher than 20%, among which 9,976 loci (99%) were polymorphic. The overall value of PIC and MAF was 0.412 and 0.227, respectively (Table 1) and 5488 out of 10 041 SNP loci (about 55%) showed a MAF value higher than 0.200 (data not shown). Descriptive statistics Calabrian cultivars analysed with the Vitis18kSNP array are summarised in Table 2. The unknown cultivars were excluded because they are not clearly assigned to some studied cultivars after cluster analysis. The average of the call rate was 0.960 and the overall observed (H_o) and expected (H_e) heterozygosity values were, respectively, 0.353 and 0.187, while the inbreeding coefficient (*F*) was -0.181 (Table 2). Taking into account the cultivars with several analysed samples > 1, genotypes belonging to the same cultivar showed frequently identical SNP profiles (Bianco d'Alessano, Gaglioppo, Guarnaccia, Magliocco Canino and Mennella Nera). SNP profile divergences among plants of the same cultivars were also observed ranged from 2 (Toccarino) to 1638 (Pecorello) loci (Table S5).

To provide greater resolution regarding the genetic background of grapevine germplasm from Calabria, we also assessed genetic evidence for signs of recent inbreeding by runs of homozygosity (ROH) detection within each cultivar in regions of homologous chromosomes that are identical by descent. The SNP panel data revealed that 14 out of 21 cultivars (67%) displayed at least one ROH between 2–13 Mb in length, with a proportion of homozygous sites ranging from 95 to 100%, the main ROH category was 3-6Mb in length (Table S6, Figure S1). The mean value of ROH in the genome of the analysed cultivars ranged from 0.625% (Prunesta) to 1.451% (Petrera) Table S7). Among others, the most homozygous cultivars were Petrera, Montonico Bianco Italico and Nzolia with an average of 7.2 Mb, 5.7 Mb and 5.5 Mb, respectively. Runs of homozygosity have been found in all chromosomes, except #19, and the number of ROHs per chromosome was greatest in #17 (4). The region including ROHs was greatest in chromosomes 8, 9, 13 and 18, which showed the largest extent of overlapping ROH positions among the analysed samples (Table S7). Interestingly, all samples belonging to the cultivar Greco Nero harboured a ROH segment on chromosome 18 involving a genomic region centred on the 1.5 Mb position, consisting in 103 SNPs and 3626.04 kb

in length (Table S8). In addition, chromosomes 6, 9, 11, 12 showed overlapping ROH regions in the samples belonging to Gaglioppo, Montonico B Italico, Nerello Mascalese, Guardavalle, respectively. Finally, chromosome 8 displayed the longest run of contiguous SNPs (355) in the samples of Montonico B Italico.

Genetic relationships, multivariate and population structure analyses

To investigate the genetic relationships among cultivars based on SNP-data, cluster analysis and kinship estimation were carried out. Many genetic distances and hierarchical clustering algorithms can be used to build trees but the most appropriate to describe the genetic diversity for a given dataset was evaluated comparing the NJ and UPGMA algorithms. The biplots and correlation indices clearly showed that the NJ algorithm defined the best representation of the original distance matrix (Figure S2). Cluster analysis using NJ algorithm and Nei (1978) distances generated a dendrogram underlining four main large clusters (Figure 2a). All samples were assigned properly to their own putative cultivar (based on sample name, origin and SSR profile); private clades for each variety were identified. The bootstraps among cluster ranged from 95 to 99% for the most nodes avoiding misclassifications (data not shown).

In parallel, a compressed mixed linear model (CMLM) was applied to study the links between marker profiles and cultivars in order to improve the ability to detect phenotype–genotype associations in the presence of population stratification and multiple levels of relatedness and to increase the statistical significance of the analysis. The kinship matrix was calculated based on the proportion of shared alleles, showing the clustering of cultivars and the dissimilarity among samples (Figure 2b). Kinship analysis confirmed the presence of four main genetic groups already suggested by cluster analysis (Figure 2a) and private branches for each investigated cultivar, of which Greco Nero was the cultivar most distant (Figure 2a,b).

The differences among cultivars was further supported by the fastSTRUCTURE algorithm. Indeed, the optimum number of genetic groups (*K*) within the Calabrian germplasm collection was determined to be *K*= 7 with a samples' clustering strongly correlated with their group membership according to fastSTRUCTURE at *K* = 7 (Figure 3). Using a cut-off of 70% ancestry, 51 out of 72 genotypes (about 70%) were classified into one of the seven major groups, in particular Greco di Bianco, Nerello Mascalese, Greco Nero showed 100% of membership for the group, red, green and blue, respectively (Figure 3). In addition, Toccarino and Castiglione belong to the purple pool (100%), like Guardavalle and Pedilongo to yellow pool, Iuvarello and Minnella nera to cyan pool.

The remaining samples, mostly belonging to Mantonico Bianco and Pecorello cultivars, showed an admixture profile without a clear assignment (Figure 3).

Principal component analysis (PCA) was applied to compare cultivars from Calabria studied here (72 genotypes) and the Sicilian cultivars investigated in Mercati et al. (2016) (Figure 4). Both regions are considered as a potential secondary centre of grapevine domestication in the Mediterranean area, and although the two regions showed F_{st} values of 0.379 (Calabria) and 0.462 (Sicily), the first and second principal components (PCs) accounted for only 8.5 and 7.7% of the variance, respectively (Figure 4). Many of the cultivars here analysed clustered together with Sicilian ones and it was confirmed when additional PCs were considered (Figure 4).

Finally, the amount and distribution of genetic diversity (Nei 1987) among Italian regions and other countries with a grapevine old tradition were investigated. The comparison study was carried out by using SSR profiles of 121 accessions/varieties from Calabria and Sicily [see Mercati et al. (2016) for the Sicilian cultivars; the accessions from Table S3 and others unpublished profiles for Calabrian cultivars/accessions] and genetic profiles already published belonging to: (i) a large panel of cultivars (120) from northern and central Italy (http://catalogoviti.politicheagricole.it/); (ii) 227 French cultivars (Lacombe et al. 2013); and (iii) 104 accessions from Spain and Portugal (Lacombe et al. 2013). A decrease of gene diversity (Nei 1987) was shown moving from Southern Italy to Europe (H_e = 0.831 and H_e = 0.729 for Sicily + Calabria and France, respectively; see also Table S9).

Parentage analysis

The SNP dataset and the probability to have identity-by-descent (IBD) alleles were investigated to assign the properly relationship category (parentage), such as parent-offspring and second degree, among cultivars. We used patterns of IBD and predictions of pedigree relationships among the 74 *V. vinifera* accessions/cultivars here analysed founding that 74% of the cultivars are related to at least one other cultivar by a first-degree relationship (Table S10). The resulting relationship among the main grapevine Calabrian and the reference cultivars can be visualised in Figure 5. The parentage analysis of germplasm analysed here is likely characterised primarily (78%) by parent-offspring (PO) with other nine relationships classified as first cousin (FC), half-sibling (HS), and in the case of Toccarino/Castiglione considered as clones (TW; Z2 = 0.961) (Table S10). Interestingly, Mantonico Bianco and Pecorello, both related to Pinot Noir and Sangiovese (as mainly revealed by fastSTRUCTURE, Figure 3), showed the highest number of relationships (76%) within the analysed panel (Table S10).

Discussion

Genetic diversity

The genetic variability among native grapevine cultivars from Calabria and their relationships were investigated by both 36 ampelographic OIV descriptors, nine SSR and 18 701 SNP loci. Simple sequence repeat analysis, first adopted to classify the accessions to their own cultivar, was able to detect 29 distinguishable genetic profiles, but not able to discriminate among accessions with different names belonging to the same cultivar (synonymies, Table S4). As expected, SSR profiles were instead able to verify the cases of accessions with the same name harbouring different SSR profiles (homonymies, Table S4). The ampelographic analysis (36 OIV descriptors) was able to discriminate among cultivars and to confirm cases of synonymies and homonymies, except for the accessions belonging to cultivar Greco Nero = Magliocco Dolce, although the eight accessions belonging the cultivar Greco Nero = Magliocco Dolce clustered in distinct groups (Figure 1). Such evidence is consistent with the definition of biotype and not in contrast with their sharing the same SSR profile (This et al. 2006). The analysis of 36 OIV descriptors was highly informative for discriminating among accessions, but these markers are time-consuming and are influenced by environment (Tessier et al. 1999).

High-throughput SNP genotyping was adopted to further study the genetic diversity and the population structure of grapevine germplasm from Calabria. The Vitis18kSNP array, developed through NGS technologies, represents a powerful tool to discover genome-wide allelic variation for genetic diversity that can be added to SSR markers for cultivar or biotypes identification. After removing SNPs having a range of NC from 20 to 100%, the analysis was carried out on the amplified loci showing a GT score higher than 0.6. The Vitis18kSNP array contains about 25% of loci identified from other *Vitis* species, thus ~2% of SNP loci (411) that did not show fragment amplification appeared reasonable, as previously reported (De Lorenzis et al. 2015, Mercati et al. 2016).

The proportion of polymorphic loci was high (99% out of used loci) while, as expected, the values of heterozygosity (expected H_e and observed, H_o) were lower than those reported for southern Italy grapevine collections previously analysed by SSR markers (Carimi et al. 2010, 2011, Mercati et al. 2012, De Lorenzis et al. 2014). These results appeared consistent with SNP bi-allelic nature, and not in contrast with their higher discriminating power due to the larger number of loci analysed. Similar results were reported analysing 700 grapevine cultivars by both 22 SSRs and 384 SNPs (Emanuelli et al. 2013) or the intra-varietal analysis of Sicilian cultivars with the same 18K SNP array and 9 SSRs (Mercati et al. 2016). The MAF mean values were also comparable with those previously

reported (Lijavetzky et al. 2007, Emanuelli et al. 2013, Mercati et al. 2016). The negative value of inbreeding coefficient (*F)* was also in agreement with previous studies (De Lorenzis et al. 2015, Mercati et al. 2016), showing a slight excess of heterozygosity probably due to a prevalence of outcrossing.

The 72 accessions analysed at nine SSR loci showed 29 different genetic profiles (Table S3) and the probability of finding different plants with the same profile at all loci appears to be low (PI = 1.08×10^{-10}), therefore identical profiles/genotypes can be considered as synonyms.

Further, to evaluate the genetic diversity the inbreeding depression and the reduced fitness among offspring of related parents, was taken into account by using pedigrees. The difficulty, however, to obtain reliable pedigree information led to developed new statistical approaches. Recently, dense SNP data were used to estimate inbreeding derived from distant ancestors in outbred populations, such as grapevine, by using the proportion of detected ROH (McQuillan et al. 2008, Keller et al. 2011, Wang et al. 2017). Runs of homozygosity are stretches of repeated homozygous sequences probably reflecting segments shared identically by descent as the result of different processes (e.g. natural and human selection, consanguinity, population size reduction). Most of the grapevine accessions showed ROH but no significant differences were found across cultivars in ROH proportion ($P > 0.05$; Table S6). The main category found in our collection was 3-6 Mb (ROH longer than 1*cM*; Figure S1) suggesting a recent events of inbreeding. Indeed, if ROHs are limited to short regions (< 0.5*cM*) length of homozygosity could be associated with more ancient founder events (Wang et al. 2017), while long ROH can be associated to recent events (Szpiech et al. 2013). As expected, high negative *F* values appeared in correlation with ROH absence, as observed in Magliocco Canino, Mantonico Bianco and Toccarino. Interestingly, we also isolated regions that overlapped among different cultivars and within genotypes belonging to the same cultivar. The results from the ROH analysis highlighted that our collection had undergone a selective pressure during the human migration with possible ROH hotspots containing genes or allelic variants that needed further investigation.

Genetic relationships, population structure and parentage analysis

Cluster and kinship analyses were able to distinguish four large clusters and some subclusters in which it is possible to classify the main cultivars from Calabria. Some relic cultivars appeared closely related to important cultivars, such as Occhi di Lepre, Mantonico Pinto and Verdina, in the same cluster with Gaglioppo, Greco Nero and Guardavalle/Mantonico Nero, respectively (Figure 2a). Many other rare cultivars appeared in a group that had no relationship with the clusters where the

main cultivars are included (Figure 2a). The first, cluster I, included three cultivars, Mantonico Bianco, Magliocco Canino and Prunesta, of which the first two are most important for Calabria grapevine biodiversity. A second large cluster II subdivided in two sub-clusters with the cultivars Nerello Mascalese, Sangiovese and Gaglioppo included in the first sub-cluster, while Castiglione and Toccarino in the second one. The third large cluster III included two sub-clusters in which Greco nero, Magliocco dolce and their synonyms formed the first sub-cluster, while Montonico, Pedilongo, Guardavalle and Mantonico Nero the second one. Pecorello, Greco di Bianco, Guarnaccia, Iuvarello and Minnella Nera together with rare cultivars, such as Lacrima Bianca, Mantonico di Nigra, Uva Ruggia and Verdana, are included in different sub-clusters of cluster IV. Kinship analysis well revealed that the cultivar Greco Nero = Magliocco Dolce appeared the farthest from the other cultivars based on SNP polymorphisms.

The distribution of cultivars among clusters was verified by STRUCTURE analysis (K = 7) demonstrating a rather large genetic diversity in the analysed accessions. In particular, cluster II includes accessions with two clearly distinguished genetic structures, green and fuchsia. Nerello Mascalese, Sangiovese and Gaglioppo shared the first one, while Castiglione and Toccarino were in the second pool (Figure 3). Cluster III is characterised by the blue genetic structure shared by the cultivar Greco Nero = Magliocco Dolce with all their synonyms of which Arvino and Magliocco Tondo are the most widespread in Calabria. The other sub-clusters included in cluster III share the yellow genetic structure with the cultivars Montonico, Nzolia, Pedilongo, Guardavalle and Mantonico Nero. As expected, cluster IV appeared the most heterogeneous group because the included cultivars showed many genetic structures (Figure 3). Indeed, genotypes with admixture profile (accessions belonging to Pecorello), but also cultivars with a unique genetic structure (Greco di Bianco or Malvasia di Lipari with red pool; Iuvarello and Mennella Nera with cyan profile) or some others with the cyan genetic pool as prevalent profile (Guarnaccia and Uva Ruggia) can be found. Finally, cluster I is characterised by a cultivar (Mantonico Bianco) sharing five out of seven genetic structures while Magliocco Canino and Prunesta are characterised by the yellow genetic structure.

Furthermore, our analysis on the amount and distribution of genetic diversity among Italian regions and European Countries provided two primary insights regarding the historical spreading of grapevine cultivation in the Mediterranean area and in Europe. First, the low variance explained by the first and second PC shows that Calabrian and Sicilian germplasm diversity cannot be summarised by one or few PCs. This evidence supports the hypothesis that the two regions have always been considered a unique and significant grapevine centre of biodiversity. Second, since the domestication bottleneck is an extended process covering several millennia, like other species subject to an

expansion history, grapevine was undergoing to frequent founder effects with genetic diversity decrease correlated to the increasing distance from its centre of domestication. Indeed, we show a genetic diversity declining by moving from Sicily and Calabria to northern Italy and Europe (0.831, 0.819 and 0.729, respectively; Table S9). This evidence could be explained by the occurrence of serial founder effects during grapevine expansion from ancestral pools, leading to a gradual increase in genetic drift and a related decrease of genetic diversity in the successive human migration and colonisation processes. In addition, the genetic similarity that populations of *V. vinifera* ssp. *sylvestris* share with cultivars in southern Italy (Garfi et al. 2013) appeared in contrast with the morphological distinction and the genetic divergence between domesticated and wild accessions (Zecca et al. 2010). Overall, our results assigned to Calabria and Sicily a key role in the grapevine domestication process, in agreement with the hypothesis that the southern Italian regions are a potential secondary centre of domestication in the Mediterranean Basin (Grassi et al. 2003, Arroyo-Garcia et al. 2006, De Mattia et al. 2008).

The grapevine is a species vegetatively propagated for thousands of years and the possible events of selfing and crosses across different generations makes an accurate pedigree reconstruction difficult starting from genomic data. Based on the large SNP dataset here developed, however, IBD analysis allowed proper relationship categories to be assigned among accessions under study, revealing closed relationships among many Calabrian cultivars. Forty-two relationships were found in our dataset, 32 of which were parent-offspring (Table S10, Figure 5). Among them, Mantonico Bianco and Pecorello showed a central role in the genetic contribution to many grapevine native cultivars from Calabria. Significant relationships between Mantonico Bianco and Guardavalle, Mantonico Bianco and Gaglioppo, Mantonico Bianco and Nerello Mascalese recently reported by using SSR markers were confirmed (Gasparro et al. 2013, Crespan et al. 2017), while the potential key role of Pecorello is reported for the first time. Such evidence is in agreement with the absence of ROHs in Pecorello (except the accession Greco di Rogliano) and Mantonico Bianco, suggesting them as potential donors of genetic variability to other cultivars. Although full parentages have not been resolved, first-degree relationships here highlighted are of interest to know the kin groups of cultivars (Vouillamoz et al. 2003), providing precious information of potential key genitors, which should be integrated with additional markers and supplementary genotypes.

Cultivar classification, synonyms and homonyms

The ancient literature available for Calabria referred to many different names frequently utilised for the same cultivar (synonymies) or less frequently different cultivars named in the same manner

(homonymies) (Table S4). In Calabria, more than in other Mediterranean regions (De Lorenzis et al. 2014, Mercati et al. 2016), the incorrect attribution of names to a specific cultivar became circular, therefore it is frequent to find two different cultivars both named with two different names depending on the sampling area. The platform of Calabrian grapevine cultivars became an inextricable interlacing of synonyms and homonyms, because more than two cultivars appeared included in this circular scheme. Among 29 genetic profiles accounted, 15 cultivars originating from Calabria and still present in cultivation were identified, of which one group of accessions appeared as synonyms (Magliocco Dolce = Greco Nero = Arvino) and two homonymies of cultivars from other Italian regions (Guarnaccia found in Calabria also as Guarnaccino and Prunesta found in Sicily as Prunestra). Another six cultivars are unknown or limited to small germplasm collection (Occhi di Lepre, Mantonico Pinto, Verdina, Lacrima Bianca, Mantonico di Nigra and Verdana). Furthermore, six cultivars frequently cultivated in Calabria were synonyms of Sicilian cultivars (Nerello Mascalese = many synonyms where Nerello is most spread, Greco di Bianco = Malvasia di Lipari, Nzolia = Inzolia = Ansonica, Petrera, Alicante di Egua = Alicante and Minnella nera). The most important cultivars from Sicilia cultivated in Calabria are Nerello Mascalese, found in several areas with names such as Nerello or Negrello, and Malvasia di Lipari traditionally cultivated in the district of Bianco (RC) where it is named 'Greco di Bianco' and from which a well known raisin wine is obtained. This synonymy determines a first confusion in the cultivar nomenclature because of another important white Calabrian cultivar (Guardavalle) is frequently found as Greco Bianco.

Finally, Iuvariello or Vuoino are synonyms of Bianco d'Alessano (from Apulia), while Montonico Bianco B (spread widely in central Italy) is frequently present in many areas of cultivation in Calabria with different names such as Montonico B Italico. Among the 15 cultivars of presumable origin from Calabria, a brief history of the most important is reported below.

The most important black skinned berry cultivars are Gaglioppo, Magliocco Dolce, Greco Nero, Magliocco Canino and Castiglione. *Gaglioppo* is the most wide spread cultivar in Calabria with 3600 ha. The name Gaglioppo, as grapevine cultivated mainly in the area of Cirò (Crotone – KR), has been reported since the second part of 18th century (Grimaldi 1770). The cultivar is registered as Gaglioppo N at the Registro Nazionale Varietà Vino (RNVV), described by Mazzei and Zappalà (1964a). The genetic stability of the cultivar appears ensured by the unique profiles founded utilising both SSR and SNP markers.

Greco Nero = Magliocco Dolce is the second most important cultivar in Calabria after Gaglioppo with 1600 ha. cultivated around the Region. It is registered at RNVV as Greco Nero N, described by Mazzei and Zappalà (1964b). Its origin is unknown but a grape named 'Magliocca' was

referred by Marafioti (1601), while ‘Lacrima’ or ‘Lagrima’ and ‘Arvino’ are its synonyms cited by Pasquale (1863) and Pagano (1901), respectively. In Calabria, the cultivar is widespread so much that it enjoys a high number of synonyms. Comparing the ampelographic schedules, Greco Nero appears rather similar to ‘Nocera’ and ‘Castiglione’ described by Mazzei and Zappalà in 1962 and 1965, respectively. Our observations confirmed that all the accessions included in the ‘group’ Greco Nero = Magliocco Dolce harbored the same SSR profile (Table S3), thus, based also on the ampelographic analysis, the accessions collected with the name Arvino, Magliocco Dolce, Magliocco Tondo, Greco N di Sibari, Mangiaguerra, are considered here as a unique cultivar (Tables S1,S2). In contrast, SNP analysis highlighted a limited variability among the accessions due to 18 SNPs describing a small sub-cluster in which Greco N di Sibari, Arvino and Mangiaguerra are included. The SSR profile harbored by the Greco Nero ‘group’ is the same reported for cultivar Castiglione N (described below) in the RNVV (http://catalogoviti.politicheagricole.it/) that in turn evidenced a different SSR profile in all the accessions analysed in this work (Table S3).

Castiglione is registered at the RNVV as Castiglione N, its origin is unknown but the first reference related to this cultivar name in Calabria is due to Grimaldi (1770), followed by Pasquale (1863). Because of the high number of synonymies, a circular mismatching of names and genotypes among Greco Nero, Magliocco Dolce and Castiglione was ascertained. In the area of Locride in the Province of Reggio Calabria, the name ‘Castiglione’ is incorrectly attributed to ‘Magliocco Dolce’ that as previously reported appeared as a synonym of Greco Nero. Other synonyms were found by Schneider et al. (2008, 2009) and then by Mannini et al. (2012).

Toccarino is another black berry cultivar from Calabria of unknown origin, unregistered at RNVV. In the past, the cultivar was probably cultivated in many areas of Calabria since the 17th century, indeed the first report was from Marafioti (1601) that referred to ‘Coccarina’. At present, Toccarino is present in old vineyards from the Tirrenian Coast in Province of Cosenza (Longobardi) to the southern part of the same Coast [Motta S. Giovanni (Province of Regio Calabria)] and the area of Locride (Ionian Coast in the Province of Reggio Calabria). Other observations on Toccarino were reported by Pasquale (1863) that referred to a grapevine named ‘Coccarino’ with deeply lobed leaves and sweet round berries by Mendola (1868) and Di Rovasenda (1877).

The presence of *Mantonico Nero* in Calabria is not completely ascertained from the ancient literature, and a cultivar with this name is not registered at RNVV. Therefore, Greco Nero, Magliocco Dolce and Castiglione were frequently named with this name as previously reported, indicating that a cultivar with black berry with this name could be present in cultivation probably in the Province of Cosenza as reported by Fera and Frojo (1881). Recently an ampelographic profile for Mantonico or

Montonico Nero was reported by Antonacci and Placco (1988), and it appeared distinguished from Mantonico Bianco, Mantonicone and Mantonico di Nigra. More recently, a certain and unique SSR profile was reported for Mantonico Nero by Crespan et al. (2017). The accession of Mantonico Nero, collected by the University of Milan described in this work, is has an ampelographic profile comparable to that described in Antonacci and Placco (1988), harboring the same SSR profile reported by Crespan et al. (2017).

The presence of *Pedilongo* in Calabria is known since the end of 19th century, cited in Province of Cosenza as Piede Longo or Coda di Volpe (Fera and Frojo 1881). Pedilongo is another important black berry cultivar from Calabria not yet registered in the RNVV. It had to be largely cultivated in the past, while nowadays the cultivar Pedilongo is present in old vineyards of both Tirrenian and Ionian southern coasts. The ampelographic profile confirmed those described by Raimondi et al. (2008).

Magliocco Canino in Calabria has been known only since the end of the 19th century (De Bonis 2002). This cultivar is registered at RNVV and it is clearly distinguishable from Magliocco Dolce for the shape of berry as reported in Calò et al. (2006) and Schneider et al. (2008). The SSR and SNP profiles are distinctive, showing a closed relationship with Mantonico Bianco sharing the same cluster (Figure 2). The ampelographic profile verified those reported in Raimondi et al. (2008).

In the areas of our sampling eight accessions resulted in synonyms of *Nerello mascalese* (see Table S3 for SSR profile), an ancient Sicilian cultivar noted by Romans (Gagliano 2008); it belongs to the group of 'Nigrelli' described by Sestini (1812). Some other black berry skin cultivars were rarely found in cultivation as *Mennella Nera*, *Prunesta*, *Mantonico di Nigra*, *Alicante* and *Petrera*. The first two are names of cultivars already reported in Sicily (De Lorenzis et al. 2014); the last one is also present in different germplasm collections in Sicily. Comparing SSR (De Lorenzis et al. 2014) or SNP (data not shown) profiles, Mennella Nera and Petrera are synonyms of those cultivars from Sicily, while Prunesta and Alicante showed divergent SSR profiles compared to that of Prunestra found in Sicily and Alicante cultivated in many Italian regions (homonymies). Finally, Mantonico di Nigra contained an unknown SSR profile compared to the main public *Vitis* databases. Another unknown SSR profile is contained by the unique red skin berry cultivar found in Calabria named *Uva Ruggia* or *Lampazona*, already cited by Raimondi et al. (2008), but for which the SSR profile is not reported.

Among the white skin berry cultivars from Calabria, Mantonico Bianco, Guardavalle and Pecorello appeared the most important. Marafioti (1601) reported the first report on a cultivated grape cultivar named *Mantonico Bianco* in the southern Ionian coast. Further information derived by

Pugliese (1849), which cited a Mantonico among white grapes for must in the area of Cirò and by Pasquale (1863), mentioning a 'Mantonica' as a white grape in Calabria. Throughout the 20th century, notices about 'Mantonico' grape cultivated in the southern Ionian coast become increasingly widespread and profuse. Recently, 'Mantonico Bianco' was registered at the RNVV (2014), clearly distinguished from 'Montonico' (found in Calabria also as 'Montonico Bianco Italico') described by Bruni (1962), both for ampelographic descriptors (Cappelleri et al. 1987) and SSR genetic profile (Crespan et al. 2008, 2017). Two biotypes of 'Mantonico' different for some morphological and oenological traits were previously described (Zappia et al. 2007), one named 'Mantonico Vera' which means 'true Mantonico' (Raimondi et al. 2008) and 'Mantonico Pizzutella' which is reported to contain the same SSR profile of 'Mantonico Vera' (Pellerone et al. 2001). In the present manuscript, 'Mantonico Bianco' cultivar was confirmed to contain a unique SSR profile (Table S3), OIV descriptors distinguishable from the other cultivars analysed (Figure 1, Table S2); in addition, a unique SNP profile was found (Figure 2). In the chapter of 'Mantonico grapes' in this and previous studies, two other SSR profiles and relic cultivars were included. '*Mantonicone*' is a white berry grape described by Crespan et al. (2008, 2017), also present in the collection Vitis International Variety Catalogue (VIVC) (www.vivc.de); in this study, this relic cultivar was collected and characterised as '*Occhi di Lepre*' as also reported by Schneider et al. (2008). '*Mantonico Pinto*', despite the adjective 'pinto' could means 'coloured', is another white berry cultivar recently reported by Crespan et al. (2017) that contains the same SSR profile described in the present study.

The grapevine cultivar *Guardavalle* has been registered at RNVV since 1971 (with its official synonym Uva Greca), more recently described by Calò et al. (2006). Its name appears derived from the homonym town in Province of Catanzaro (Pagano 1901). Among its synonyms, most frequent is 'Uva da Passito' (Province of Kroton) which means 'grapes to dry out' and recalls the historical name 'Passulara' for the good aptitude to grape drying of this cultivar (Schneider et al. 2008, 2009). More recently, Crespan et al. (2017) described some biotypes named 'Montonico di Rogliano', a small town in Province of Cosenza as another synonym as well as 'Greco Bianco' utilised mainly in the area of Cirò. In the RNVV, in addition to Guardavalle B there are two other cultivars registered, Greco B and Greco Bianco B, the first is a distinct genotype from Campania also known as 'Greco di Tufo', while the second is a duplicate of Guardavalle B (Schneider et al. 2013).

Until a few years ago, the cultivar *Pecorello* was considered a relic; the white berry grapevine of unknown origin was rarely present in old vineyards of Savuto Valley area (Province of Cosenza), where it is improperly exchanged with the cultivar 'Pecorino', widely spread in central Italy. In the

last decade, the interest in Pecorello grapevine cultivation has increased, especially in the DOC area Donnici in Province of Cosenza.

Pecorello is also registered at RNVV without any official synonyms, but in some areas is named ‘Greco Bianco’ or ‘Greco Bianco di Rogliano’, synonyms to avoid for the presence in the same area of Guardavalle which the official synonym is ‘Uva Greca’ frequently named also ‘Greco Bianco’ as previously reported (Crespan et al. 2017). The SSR profile of Pecorello and the ampelographic descriptors reported by Costacurta et al. (2004) are the same of those reported in the RNVV (Calò et al. 2006) and finally matching to those previously reported for ‘Greco Bianco di Rogliano’ by Antonacci and Placco (1997). Instead, the SNP data here reported appeared to underline the presence of two largely different Pecorello, one of which is named ‘Greco di Rogliano’. Because of the high number, however, of polymorphic SNP loci between the sample named Greco di Rogliano and the others belonging to Pecorello (Table S4), probably a misclassification occurred during the sampling.

Another important white berry cultivar cultivated in the southern Ionian coast of Calabria is *Malvasia di Lipari*, in the town of Bianco. In this area, it is named erroneously ‘Greco di Bianco’ that is not reported as official synonym in the RNVV where the cultivar was registered as Malvasia di Lipari (Mazzei and Zappalà 1964c). Historical writings that testify to the presence in Calabria of this grapevine date from around 1600, where in a list of 19 grapevines both ‘Greco’ and ‘Greca’ were reported (Marafioti, 1601). More recently, in a report of the economic and agricultural status of Calabria, a ‘Greco del Tirreno’ and a ‘Greca di Gerace’ (Gerace is a town close to Bianco) were cited, distinguishing them (Pasquale 1863). The author claims that the ‘Greca’ grape is sweet, more than ‘Greco’, probably referred to *Guardavalle* or *Greco Bianco* described above. *Greco di Bianco* alias Malvasia di Lipari is close also to ‘Malvasia di Sardegna\’ containing the same SSR profile, although two different cultivars were registered in the RNVV during ’70 years of last century. The same SSR profile was detected also for ‘Malvasia de Sitges’ and ‘Malvasia dubrovacka’ in Spain and Croatia (Crespan et al. 2006), as well as for ‘Malvasia Candida in Madeira and ‘Malvasia de La Palma’ in the Canary islands (Rodrìguez-Torres et al. 2009). The presence of *Malvasia di Lipari* in Sicily has also been known since the 17th century (Cupani 1696). As reported for Calabria, however, it was probably introduced into Sicily during the 6th century BC by the Greeks (Bica 2007). Malvasia is cultivated in the provinces of Messina, in particular in volcanic Aeolian Isles (Hammer and Laghetti 2006).

Finally, in the collection described here an accession named ‘*Nzolia*’ is included and which is frequently found in cultivation in Calabria with other names, synonym of *Inzolia*, the ancient Sicilian cultivar noted by Roman authors such as Pliny (Gagliano 2008). We also found two accessions with

the name '*Iuvarello*' and '*Vuiono*' that are the most frequent synonyms found in Calabria for another important white berry cultivar from Apulia, namely *Bianco d'Alessano*. This white berry grape has an uncertain origin, albeit its origin has been identified in the Valle d'Itria (Province of Taranto, Apulia) (Di Rovasenda 1877). Two other accessions '*Guarnaccia*' and '*Lacrima bianca*' appeared as homonyms of *Guarnaccia B* (registered at RNVV) and *Lacrima bianca* (present in the Italian Vitis Database), while '*Verdana*' and '*Verdina*', frequently found in old vineyards of the Province of Cosenza with similar names, showed unknown and different SSR profiles.

Conclusion

Large grapevine repositories are essential investigation tools for studying genetic diversity, relationships and parentage analysis. Such single collections are not usually exhaustive for detecting all the spread and relic cultivars present in an area of cultivation, although we spent a large effort to collect a representative platform of grape genetic diversity available in Calabria.

The assignment of a true-to-type name to each accession/cultivar is essential for the identification of synonyms and homonyms. Then, a panel of OIV descriptors together with selected molecular markers for determining morphological and genetic profiles remained the first goal of this type of work, a first milestone for further successful determination of genetic relationships.

We report the first significant report on the assessment of genetic diversity of the cultivated grapevine platform from Calabria in southern Italy. The importance of this region is indicated by many historical and archaeological documents that indicate *Enotria* as a possible secondary centre of grape domestication. The main OIV descriptors, the standard SSR panel and the Vitis18kSNP array allowed a description of a high genetic variability and to recognise the true-to-type cultivars of this area detecting many synonyms. Our dataset was also useful to underline an admixture genetic structure and many parent-offspring relationships between cultivars, highlighting a central role of Mantonico Bianco and Pecorello in the grapevine pedigree of southern Italy. Most interestingly, the genetic diversity found in southern Italy decreased passing from northern Italy to European grape cultivars. In summary, we provide a first evidence of a central role of *Enotria* in the grape genetic diversification of Mediterranean Basin.

Acknowledgements

This work was partially funded by Agroalimentare e Ricerca (AG.E.R.) grant n. 2010-2104 An Italian Vitis database with multidisciplinary approach, for exploitation and valorization of the regional

genotypes. Further, several grants funded by Regione Calabria were utilised to collect and maintain the collection for obtaining all the data included in this manuscript.

References

Alleweldt, G. and Possingham, J.V. (1988) Progress in grapevine breeding. Theoretical and Applied Genetics **75**, 669-673.

Antonacci, D. and Placco, L. (1988) Mantonico. Regione Calabria - assessorato agricoltura and istituto sperimentale per la viticoltura (sezione operativa). Bari, E.D.L., ed. Mapograf: Vibo Valentia, Italy.

Antonacci, D. and Placco, L. (1997) Vitigni di qualità di antica coltivazione nell'Italia meridionale: i vitigni Greci nella viticoltura calabrese. Vignevini **7/ 8,** 3 -23.

Arroyo-Garcìa, R., Ruiz-Garcìa, L., Bolling, L., Ocete, R., Lòpez, M.A., Arnold, C., Ergul, A., Söylemezoğlu, G., Uzun, H.I., Cabello, F., Ibáñez, J., Aradhya, M.K., Atanassov, A., Atanassov, I., Balint, S., Cenis, J.L., Costantini, L., Goris-Lavets, S., Grando, M.S., Klein, B.Y., McGovern, P.E., Merdinoglu, D., Pejic, I., Pelsy, F., Primikirios, N., Risovannaya, V., Roubelakis-Angelakis, K.A., Snoussi, H., Sotiri, P., Tamhankar, S., This, P., Troshin, L., Malpica, J.M., Lefort, F. and Martinez-Zapater, J.M. (2006) Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *sativa*) based on chloroplast DNA polymorphisms. Molecular Ecology **15,** 3707-3714.

Barkley, N.A., Roose, M.L., Krueger, R.R. and Federici, C.T. (2006) Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRs). Theoretical and Applied Genetics **112,** 1519-1530.

Bevilacqua, P. (1980) Le campagne del Mezzogiorno tra fascismo e dopoguerra. Il caso della Calabria. Einaudi: Torino, Italy.

Biagini B., De Lorenzis, G., Imazio, S., Failla, O. and Scienza, A. (2014) Italian wild grapevine (*Vitis vinifera* L. subsp. *sylvestris*) population: insights into eco-geographical aspects and genetic structure. Tree Genetics & Genomes **10,** 1369-1385.

Bica, D. (2007) Vitigni di Sicilia (Priulla Publisher: Palermo, Italy).

Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics **32,** 314-331.

Bowers, J.E., Dangl, G.S. and Meredith, C.P. (1999) Development and characterization of additional microsatellite DNA markers for grape. American Journal of Enology and Viticulture **50,** 243-246.

Bowers, J.E., Dangl, G.S., Vignani, R. and Meredith, C.P. (1996) Isolation and characterization of new polymorphic simple sequence repeat loci in grape (*Vitis vinifera* L.). Genome **39**, 628-633.

Brun, J.P. and Laubenheimer, F. (2001) La viticulture en Gaule. Gallia **58,** 203–219.

Bruni, B. (1962) Mantonico bianco. Principali Vitigni da Vino coltivati in Italia. Vol. III. (Ministero dell'Agricoltura e delle Foreste: Rome, Italy).

Buxò, R. (2008) The agricultural consequences of colonial contacts on the Iberian Peninsula in the first millennium BC. Vegetation History and Archaeobotany **17,** 145–154.

Cabezas, J.A., Ibáñez, J., Lijavetzky, D., Vélez, D., Bravo, G., Rodríguez, V., Carreño, I. Jermakow, A.M., Carreño, J., Ruiz-García, L., Thomas, M.R. and Martinez-Zapater, J.M. (2011) A 48 SNP set for grapevine cultivar identification. BMC Plant Biology **11,** 153.

Calò, A., Scienza, A. and Costacurta, A. (2006) Vitigni d'Italia (Edagricole: Bologna, Italy).

Cappelleri, G., Lovino, R. and Ambrogio, R. (1987) Sull'idoneità di alcune varietà di uva meridionali a produrre vini "moderni". Vignevini **5,** 13-18.

Carimi, F., Mercati, F., Abbate, L. and Sunseri, F. (2010) Microsatellite analyses for evaluation of genetic diversity among Sicilian grapevine cultivars. Genetic Resources and Crop Evolution **57,** 703-719.

Carimi, F., Mercati, F., Michele, R., Fiore, M.C., Riccardi, P. and Sunseri, F. (2011) Intra-varietal genetic diversity of the grapevine (*Vitis vinifera* L.) cultivar "Nero d'Avola" as revealed by microsatellite markers. Genetic Resources and Crop Evolution **58**, 967–975.

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

Cervera, M.T., Cabezas, J.A., Sancha, J.C., Martinez de Toda, F. and Martinez-Zapater J.M. (1988) Application of AFLPs to the characterization of grapevine *Vitis vinifera* L. genetic resources. A case study with accessions from Rioja (Spain). Theoretical and Applied Genetics **97,** 51-59.

Cipriani, G., Spadotto, A., Jurman, I., Gaspero, G., Crespan, M., Meneghetti, S., Frare, E. Vignani, R., Cresti, M., Morgante, M., Pezzotti, M., Pe, E., Policriti, A. and 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. Theoretical and Applied Genetics **121,** 1569-1585.

Costacurta, A., Calò, A., Antonacci, D., Catalano, D., Crespan, M., Cararro, R., Giust, M., Agio, L., Ostan, M., Di Stefano, R. and Borsa, D. (2004) La caratterizzazione dei Greci e Grechetti a bacca bianca coltivati in Italia. Rivista di Viticoltura e Enologia **58**, 3 -20.

Crespan, M., Cabello, F., Giannetto, S., Ibanez, J., Karoglan Kontic, J., Maletic, E., Pejic, I., Rodrìgez-Torres, I. and Antonacci, D. (2006) Malvasia delle Lipari, Malvasia di Sardegna, Greco

di Gerace, Malvasia de Sitges and Malvasia dubrovačka – synonyms of an old and famous grape cultivar. Vitis **45**, 69–73.

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

Crespan, M., Storchi, P. and Migliaro D. (2017) Mantonico bianco grapevine cultivar is the second parent of the Sicilian Catarratto. American Journal Enology and Viticulture **68,** 258-262.

Cupani, F. (1696) Hortus Catholicus seù illustrissimi, & excellentissimi principis Catholicae (Ed. Benzi: Napoli, Italy).

De Bonis, M. (2002) Terra d'uve. Vini e vitigni in Calabria dall'antichità all'Ottocento. (Ed. Le nuvole: Cosenza, Italy).

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

De Lorenzis, G., Chipashvili, R., Failla, O. and 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 Biology **15**, 154.

De Mattia, F., Imazio, S., Grassi, F., Baneh, H.D., Scienza, A. and Labra M. (2008) Study of Genetic Relationships Between Wild and Domesticated Grapevine Distributed from Middle East Regions to European Countries. Rendiconti Lincei **19,** 223 -240.

Di Rovasenda, G. (1877) Saggio di una ampelografia universale (Loescher: Torino, Italy).

Di Vora, A., and Castelletti L. (1995) Indagine preliminare sull'archeologia della vite (*Vitis vinifera* L.) in base ai caratteri diagnostici del vinacciolo. Rivista Archeologica dell'Antica Provincia e Diocesi di Como **176,** 333–358.

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

Fera, M., and Frojo, G. (1881) Primi studi ampelografici della Provincia di Cosenza. Bollettino ampelografico, Fasc. XV, 163-186 (Ministero di Agricoltura, Industria e Commercio: Napoli, Italy).

Forni, G. (2012) The origin of “Old World” viticulture. Maghradze, D., L. Rustioni, A. Scienza, J. Turok, O., eds. Caucasus and Northern Black Sea Region. Volume 1. 1st ed. Failla. Vitis Special Issue, 27–38.

Gagliano, F. (2008) I vitigni autoctoni siciliani. Viticultura ed enologia di Sicilia. Dalla ricerca in vigna e in cantina alle strategie di marketing (Azienda Grafica Editoriale: Catania, Italy).

Galet, P. (1979) A practical ampelography: grapevine identification (University Press: Ithaca, NY, USA).

Galet, P. (1991) Prècis d’ampèlographie pratique. Galet, P., ed. 6 ed (Imprimerie Dèhan. Montpellier, France).

Garfì, G., Mercati, F., Fontana, I., Collesano, G., Pasta, S., Vendramin, G.G., De Michele, R. and Carimi, F. (2013) Habitat features and genetic integrity of wild grapevine *Vitis vinifera* L. subsp. *sylvestris* (C.C. Gmel.) Hegi populations: a case study from Sicily. Flora **208,** 538– 548.

Gasparro, M., Caputo, A.R., Bergamini, C., Crupi, P., Cardone, M.F., Perniola, R. and Antonacci, D. (2013) Sangiovese and its offspring in Southern Italy. Molecular Biotechnology **54,** 581-589.

González-Techera, A., Jubany, S., Ponce de León, I., Boido, E., Dellacassa, E., Carrau, F.M., Hinrichsen, P., and Gaggero, C. (2004) Molecular diversity within clones of cv. Tannat (*Vitis vinífera*). Vitis **43**, 179-185.

Goudet, J. (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. Molecular Ecology Notes **5,** 184-186.

Grassi, F., Labra, M., Imazio, S., Spada, A., Sgorbati, S., Scienza, A. and Sala, F. (2003) Evidence of a secondary grapevine domestication centre detected by SSR analysis. Theoretical and Applied Genetics **107,** 1315-1320.

Grimaldi, D. (1770) Saggio di economia campestre per la Calabria ultra (Presso Vincenzo Orsini:, Napoli, Italy).

Gu, Z., Gu, L., Eils, R., Schlesner, M. and Brors, B. (2014) Circlize implements and enhances circular visualization in R. Bioinformatics **30**, 2811-2812.

Hammer, K. and Laghetti, G. (2006) Small agricultural islands and plant genetic resources. (Ed. Istituto Genetica Vegetale (IGV) - Consiglio Nazionale delle Ricerche (CNR): Bari, Italy.

IPGRI, UPOV, OIV (1997) Descriptors for Grapevine (*Vitis* spp.). International Union for the Protection of New Varieties of Plants, Geneva, Switzerland/Office International de la Vigne et du Vin, Paris, France/International Plant Genetic Resources Institute, Rome, Italy.

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., Bruyère, 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., Pè, M.E., Valle, G., Morgante, M., Caboche, M., Adam-Blondon, A.F., Weissembach, J., Quétier, F., Wincker, P. and French–Italian Public Consortium for Grapevine genome Characterization. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449,** 463-467.

Jombart, T. and Ahmed, I. (2011) adegenet 1.3-1: new tools for the analysis of genome-wideSNP data. Bioinformatics Applications Note **27**, 3070-3071.

Keller, M.C., Visscher, P.M., Goddard, M.E. and Rosenberg, N.A. (2011) Quantification of Inbreeding Due to Distant Ancestors and Its Detection Using Dense Single Nucleotide Polymorphism Data. Genetics **189**, 237-249.

Kurokawa, Y., Nodaa, T., Yamagatac, Y., Angeles-Shima, R., Sunoharaf, H., Ueharaa, K., Furuta, T., Nagai, K., Jena, K.K., Yasui, H., Yoshimura, A., Ashikari, M. and Doi, K. (2016) Construction of a versatile SNP array for pyramiding useful genes of rice. Plant Science **242,** 131–139.

Lacombe, T., Boursiquot, J.M., Laucou, V., Di Vecchi-Staraz, M. Péros, J.P. and This, P. (2013) Large-scale parentage analysis in an extended set of grapevine cultivars (*Vitis vinifera* L.). Theoretical and Applied Genetics **126**, 401-414.

Laucou, V., Lacombe, T., Dechesne, F., Siret, R., Bruno, J.P., Dessup, M., Dessup, T., Ortigosa, P., Parra, P., Roux, C., Santoni, S., Varès, D., Péros, J.P., Boursiquot, J.M. and This, P. (2011) High throughput analysis of grape genetic diversity as a tool for germplasm collection management. Theoretical and Applied Genetics **122,** 1233-1245.

Laidò, G., Mangini, G., Taranto, F., Gadaleta, A., Blanco, A., Cattivelli, L., Marone, D., Mastrangelo, A.M., Papa, R. and 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**, e67280.

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, J.M., Morgante, M, Raj, P.M., Ponnaiah, M., Quesneville, H., Scalabrin, S., Torres-Perez, R. and Adam-Blondon, A.F. (2013) The GrapeReSeq 18 k Vitis genotyping chip. Pinto, M., ed. Proceedings of the IX international symposium grapevine physiology and biotechnology; 21–26

April 2013; La Serana, Chile (International Society for Horticultural Science: Leuven, Belgium) p. 123.

Levadoux, L. (1956) Les populations sauvages et cultivées de *Vitis vinifera* L. Annales de l'Amélioration des Plantes **6**, 59-118.

Lijavetzky, D., Cabezas, J.A., Ibáñez, A., Rodríguez, V. and Martínez-Zapater, J.M. (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, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., Gore, M.A., Buckler, E.S. and Zhang, Z. (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics **28,** 2397-2399.

Magi, A., Tattini, L., Palombo, F., Benelli, M., Gialluisi, A., Giusti, B., Abbate, R., Seri, M., Gensini, G.F., Romeo, G. and Pippucci, T. (2014) H3m2: detection of runs of homozygosity from whole-exome sequencing data. Bioinformatics **30,** 2852-2859.

Mannini, F., Mollo, A., Cuozzo, D., Lanati, D. and Marchi, D. (2012) Vitigni di Calabria – Selezione e potenzialità enologiche. (Ed. Librandi: Cirò Marina, Italy).

Marafioti, G. (1601) Croniche et antichità di Calabria. (Ed. Lorenzo Pasquati ad instanza de gl'Vniti: Padova, Italy).

Maul, E., Sudharma, K.N., Kecke, S., Marx, G., Müller, C., Audeguin, L., Boselli, M., Boursiquot, J.M., Bucchetti, B., Cabello, R., Carraro, R., Crespan, M., de Andrés, M.T., Eiras Dias, J., Ekhvaia, J., Gaforio, L., Gardiman, M., 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, G., 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, S., Staraz, M., Zulini, L., Bacilieri, R. and 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.

Mazzei, A. and Zappalà, A. (1962) Nocera. Principali vitigni da vino coltivati in Italia. Vol. III (Ministero dell'Agricoltura e delle Foreste: Roma, Italy).

Mazzei, A. and Zappalà, A. (1964a) Gaglioppo. Principali vitigni da vino coltivati in Italia. Vol. III (Ministero dell'Agricoltura e delle Foreste: Roma, Italy).

Mazzei, A. and Zappalà, A. (1964b) Greco nero. Principali vitigni da vino coltivati in Italia. Vol. III (Ministero dell'Agricoltura e delle Foreste: Roma, Italy).

Mazzei, A. and Zappalà, A. (1964c) Malvasia di Lipari. In: Principali Vitigni da Vino coltivati in Italia, Vol. III, Ministero dell'Agricoltura e delle Foreste, Roma.

Mazzei, A. and Zappalà, A. (1965) Castiglione. Principali vitigni da vino coltivati in Italia. Vol. III (Ministero dell'Agricoltura e delle Foreste: Roma, Italy).

McGovern, P.E. (2003) Ancient wine: the search for the origins of viniculture (Princeton University Press: Princeton, NJ, USA)).

McQuillan, R., Leutenegger, A.L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., Macleod, A.K., Farrington, S.M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S.H., Dunlop, M.G., Wright, A.F., Campbell, H. and Wilson, J.F. (2008) Runs of homozygosity in European populations. American Journal of Human Genetics **83,** 359-372.

Melo, A.T.O., Bartaula, R. and 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.

Mendola, A. (1868) Trattato del Catalogo della collezione di viti italiane e straniere radunate in Favara. (Parrino and Carini Publishers: Favara, Italy).

Mercati, F., Riccardi, P., Leebens-Mack, J., Abenavoli, M.R., Falavigna, A. and Sunseri, F. (2013) Single nucleotide polymorphism isolated from a novel EST dataset in garden asparagus (*Asparagus officinalis* L.). Plant Science **203,** 115-123.

Mercati, F., Dattola, A., Torello Marinoni, D., Longo, C., Alba, V., Gullo, G., Zappia, R., Schneider, A. and Sunseri, F. (2012) Characterization of grapevine germplasm from South Italy regions by using microsatellites markers. Proceedings IX Convegno Nazionale della Biodiversità (ed. CIHEAM-IAMB: Valenzano, Italy).

Mercati, F., De Lorenzis, G., Brancadoro, L., Lupini, A., Abenavoli, M.R., Barbagallo, M.G., Di Lorenzo, R., Scienza, A. and Sunseri F. (2016) High-throughput 18K SNP array to assess genetic variability of the main grapevine cultivars from Sicily. Tree Genetics & Genomes **12**, 59.

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

Morgante, M. and Olivieri, A.M. (1993) PCR-amplified microsatellites as markers in plant genetics. Plant Journal **3**, 175-182.

Myles, S., Chia, J.M., Hurwitz, B., Simon, C., Zhong, G.Y., Buckler, E. and Ware, D. (2010) Rapid genomic characterization of the cenus Vitis. PLoS ONE **5**, e8219.

Myles, S., Boyko, A.R., Owens, C.L., Brown, P.J., Grassi, F., Aradhya, M.K., Prins, B., Reynolds, A., Chia, J.M., Ware, D., Bustamante, C.D. and Buckler, E.S. (2011) Genetic structure and domestication history of the grape. Proceeding of the National Academy of Science of the United States of America **108,** 3530-3535.

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

Pagano, L. (1901) Studi su la Calabria. (Ed. D'Auria: Napoli, Italy).

Park, S.D.E. (2001) Trypanotolerance in West African cattle and the population genetic effects of selection. PhD thesis. University of Dublin, Dublin, Ireland.

Pasquale, G.A. (1863) Relazione sullo stato fisico-economico-agrario della Prima Calabria Ulteriore. (Ed. R. Albergo de' Poveri: Napoli, Italy).

Pellerone, F.I., Edwards, K.J. and Thomas, M.R. (2001) Grapevine microsatellite repeats: isolation, characterisation and use for genotyping of grape germplasm from southern Italy. Vitis **40**, 179–186.

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

Pindo, M., Vezzulli, S., Coppola, G., Cartwright, D.A., Zharkikh, A., Velasco, R. and Troggio, M. (2008) SNP high-throughput screening in grapevine using the SNPlex genotyping system. BMC Plant Biology **8,** e12.

Pugliese, G.F. (1849) Descrizione ed istorica narrazione dell'origine e vicende politico-economiche di Cirò (Ed. Fibreno: Napoli, Italy).

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. and Sham, P.C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics **81,** 559-575.

Raj, A., Stephens, M. and Pritchard, J.K. (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics **197,** 573-589.

Raimondi, S., Schneider, A., De Santis, D. and Cavallo, L. (2008) Schede ampelografiche. In: AA VV - Il Gaglioppo e i suoi fratelli. I vitigni autoctoni calabresi (Tecniche Nuove: Verona, Italy) pp. 125-130.

Rivera Nunez, D. and Walker, M.J. (1989) A review of palaeobotanical findings of early *Vitis* in the Mediterranean and the origins of cultivated grape-vines, with special reference to new pointers to

prehistoric exploitation in the Western Mediterranean. Review of Palaeobotany and Palynology **61**, 205–237.

Rodríguez-Torres, I., Ibáñez, J., De Andrés, M.T., Rubio, C., Borrego, J., Cabello, F., Zerolo, J. and Muñoz-Organero, G. (2009) Synonymus and homonymus of 'Malvasía' cultivars (*Vitis vinifera* L.) existing in Spain. Spanish Journal of Agricultural Research **7,** 563-571.

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

Schneider, A., Raimondi, S., Grando, M.S., Zappia, R. De Santis, D., Torello Marinoni, D. and Librandi, N. (2008) Studi per il riordino del germoplasma viticolo della Calabria. AA VV - Il Gaglioppo e i suoi fratelli. I vitigni autoctoni calabresi (Tecniche Nuove: Verona, Italy) pp. 117-125. Schneider, A., Raimondi, S., Moreira, F.M., De Santis, D., Zappia, R., Torello Marinoni, D., Librandi, N. and Grando, M.S. (2009) Contributo all'identificazione dei principali vitigni calabresi. Frutticoltura **1-2,** 46-55.

Schneider, A. (2013) Vitigni autoctoni di Calabria. Varietà locali: districarsi tra sinonomi ed omonomi. L'Informatore Agrario **11**, 6-8.

Sefc, K.M., Regner, F., Tureschek, E., Glossl, J. and Steinkellner, H. (1999) Identification of microsatellite sequences in *Vitis riparia* and their application for genotyping of different *Vitis* species. Genome **42,** 367-373.

Sestini, D. (1812) Dei vini di mascali della Sicilia. Manuscript 12c (Accademia dei Gorgofili: Firenze, Italy)

Sim, S.C., van Deynze, A., Stoffel, K., Douches, D.S., Zarka, D., Ganal, M.W., Chetelat, R.T., Hutton, S.F., Scott, J.W., Gardner, R.G., Panthee, D.R., Mutschler, M., Myers, J.R. and Francis, D.M. (2012) High-Density SNP Genotyping of tomato (*Solanum lycopersicum* L.) reveals patterns of genetic variation due to breeding. PLoS ONE **7,** e45520.

Szpiech, Z.A., Xu, J., Pemberton, T.J., Peng, W., Zöllner, S., Rosenberg, N.A. and Li, J.Z. (2013) Long runs of homozygosity are enriched for deleterious variation. American Journal of Human Genetics **93,** 90-102.

Tessier, C., David, J., This, P., Boursiquot, J-M. and Charrier, A. (1999) Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. Theoretical and Applied Genetics **98**, 171-177.

Teti, V. (2008) Il vino nella storia calabra. AA VV - Il Gaglioppo e i suoi fratelli. I vitigni autoctoni calabresi (Ed. Tecniche Nuove: Verona, Italy) pp. 1-25.

This, P., Lacombe, T. and Thomas, M.R. (2006) Historical origins and genetic diversity of wine grapes. Trends in Genetics **22**, 511-519.

Thomas, M.R. and Scott, N.S. (1993) Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequence-tagged sites (STSs). Theoretical and Applied Genetics **86,** 985-990.

VanRaden, P.M. (2008) Efficient methods to compute genomic predictions. Journal of Dairy Science **91,** 4414-4423.

Velasco, R., Zharkikh, A., Troggio, M., Cartwright, D.A., Cestaro, A., Pruss, D., Pindo, M., Fitzgerald, L.M., Vezzulli, S., Reid, J., Malacarne, G., Iliev, D., Coppola, G., Wardell, B., Micheletti, D., Macalma, T., Facci, M., Mitchell, J.T., Perazzolli, M., Eldredge, G., Gatto, P., Oyzerski, R., Moretto, M., Gutin, N., Stefanini, M., Chen, Y., Segala, C., Davenport, C., Damattè, 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, J.A., Sterck, L., Vandepoele, K., Grando, S.M., Toppo, S., Moser, C., Lanchbury, J., Bogden, R., Skolnick, M., Sgaramella, V., Bhatnagar, S.K., Fontana, P., Gutin, A., Van de Peer, Y., Salamini, F. and 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, C.T., Gasic, K., Micheletti, D., Rosyara, U.R., Cattonaro, F., Vendramin, E., Main, D., Aramini, V., Blas, A.L., Mockler, T.C., Bryant, D.W., Wilhelm, L., Troggio, M., Sosinski, B., Aranzana, M.J., Arús, P., Iezzoni, A., Morgante, M. and 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.

Vivier, M.A. and Pretorius, I.S. (2002) Genetically tailored grapevines for the wine industry. Trends in Biotechnology **20,** 472-478.

Vouillamoz, J., Maigre, D. and Meredith, C.P. (2003) Microsatellite analysis of ancient alpine grape cultivars: pedigree reconstruction of *Vitis vinifera* L. Cornalin du Valais. Theoretical and Applied Genetics **107,** 448-454.

Wang, L., Beissinger, T.M., Lorant, A., Ross-Ibarra, C., Ross-Ibarra, J. and Hufford, M. (2017) The interplay of demography and selection during maize domestication and expansion. BioRxiv 114579.

Winfield, M.O., Allen, A.M., Burridge, A.J., Barker, G.L., Benbow, H.R., Wilkinson, P.A., Coghill, J.,Waterfall, C., Davassi, A., Scopes, G., Pirani, A., Webster, T., Brew, F., Bloor, C., King, J., West, C., Griffits, S., King, I., Bentley, A.R. and Edwards, K.J. (2015) High density SNP

genotyping array for hexaploid wheat and its secondary and tertiary gene pool. Plant Biotechnology Journal **14,** 1195-1206.

Wright, S. (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution **19,** 395-420.

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

Zappia, R., Gullo, G., Mafrica, R. and Di Lorenzo, R. (2007) Mantonico vera e Mantonico pizzutella: descrizione ampelografica, analisi microsatellite e comportamento bio-agronomico. ISSN: 1127-3496. Italus Hortus **14**, 59-62.

Zecca, G., De Mattia, F., Lovicu, G., Labra, M., Sala, F. and Grassi, F. (2010) Wild grapevine: *sylvestris*, hybrids or cultivars that escaped from vineyards? Molecular evidence in Sardinia. Plant Biology **12,** 558-562.

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

Zohary, D., Hopf, M. and Weiss, E. (2000) Domestication of plants in the Old World. 3d ed. (Oxford University Press: New Yor, NY, USA) pp. 151-159.

Zohary, D. and Hopf, M. (1993) Domestication of plants in the old world (Clarendon Press: Oxford, England) pp 143-150.

Tables

Table 1. Summary statistics of genetic variation obtained by the Vitis18kSNP array in 72 grapevine genotypes belonging to the main cultivars of Calabria and to the two reference cultivars, Pinot Noir and Sangiovese.

Parameters	Values
n†	74
Total number of loci	18 071
100% failed loci	411
GT > 0.6	10 295
No. of used loci	10 041
No. of polymorphic loci	9776
PIC	0.412
MAF	0.227
H_o‡	0.352
H_e	0.190
F§	−0.181

- † Sample size. *F*, inbreeding coefficient; H_e, expected heterozygosity; H_o, observed heterozygosity; MAF, minor allele frequency; PIC, Polymorphic information content

Table 2. Genetic diversity of Calabrian cultivars revealed by the Vitis18kSNP array.

.

Cultivar	*n*†	CR	H_o		*F*
Alicante	1	0.989	0.331	192	–
Bianco d'Alessano	2	0.984	0.308	154	−0.007
Castiglione	6	0.982	0.328	164	−0.064
Gaglioppo	4	0.987	0.341	171	−0.117
Greco Nero	11	0.984	0.330	166	−0.079
Guardavalle	4	0.987	0.300	150	0.022
Guarnaccia	2	0.987	0.325	0.163	−0.064
Inzolia	1	0.992	0.309	0.169	–
Magliocco Canino	3	0.800	0.447	0.224	−0.599
Malvasia di Lipari	3	0.986	0.360	0.180	−0.176
Mantonico Bianco	3	0.912	0.690	0.354	−0.620
Mantonico Nero	1	0.985	0.334	0.180	−0.093
Mennella Nera	2	0.983	0.303	0.152	0.007
Montonico Bianco Italico	3	0.986	0.296	0.148	0.031
Nerello Mascalese	8	0.981	0.346	0.173	−0.127

Cultivar	*n*†	CR	H_o	H_e	*F*
Parmisana	2	0.987	0.331	0.164	−0.078
Pecorello	4	0.847	0.387	0.293	−0.775
Petrera	1	0.983	0.317	0.190	–
Prunesta	2	0.983	0.332	0.161	−0.059
Toccarino	2	0.852	0.367	0.183	−0.272
Uva Ruggia	1	0.987	0.321	0.189	–
Mean	3.1	0.960	0.353	0.187	−0.181

- † Sample size. CR, call rate (out of a 10 041 used markers); H_e, expected heterozygosity; H_o, observed heterozygosity; *F*, inbreeding coefficient; −, not detected.

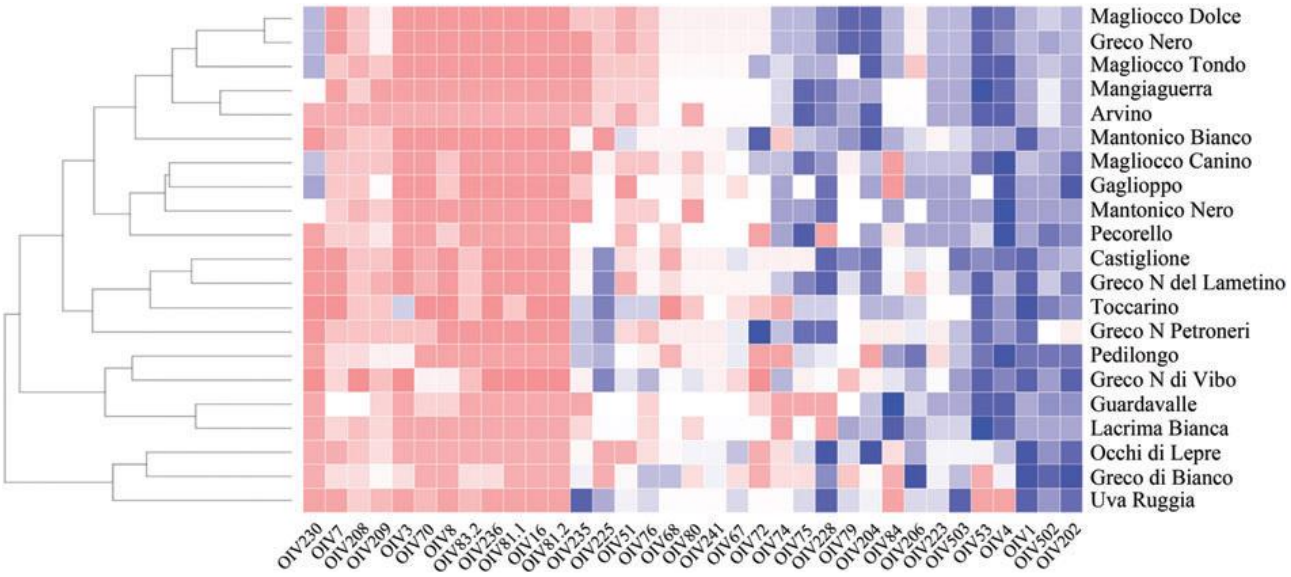

Figure 1. Thirty-six Organisation Internationale de la Vigne et du Vin (OIV) descriptors recorded for the accessions belonging to the main grapevine cultivars from Calabria. Heatmap was used to describe the variation in the OIV descriptors. Different colours and gradients represent the categories for each descriptor reported in Table S2.

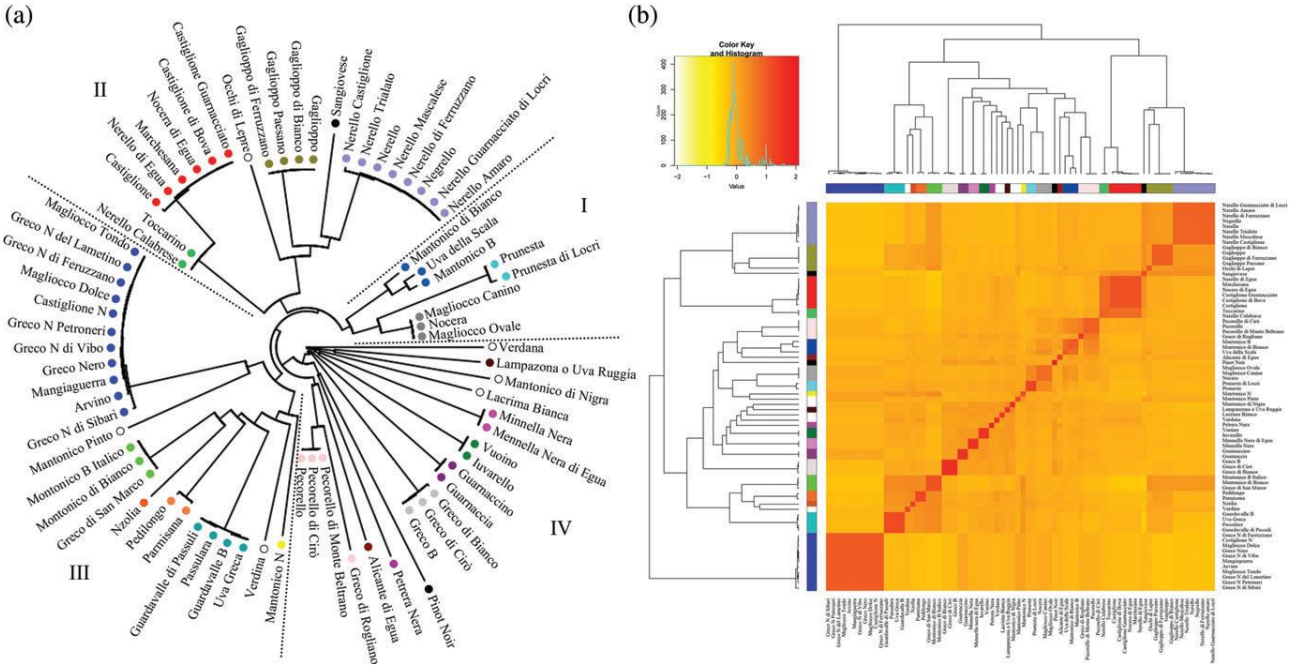

Figure 2. Genetic relationships among the 72 samples studied, belonging to the main grapevine cultivars of Calabria and to the two reference cultivars, Pinot Noir and Sangiovese, obtained by 9207 polymorphic single nucleotide polymorphisms (SNPs). (a) Dendrogram generated using the neighbourhood joining (NJ) method and Nei's distance. Dotted lines separate the four main clusters (I–IV). (b) Kinship analysis visualised through a heatmap. Colour gradient displays the dissimilarity among genotypes: red indicates the most similar clones, while white shows the lowest genetic similarity. The cultivars include Alicante (●); Bianco d'Alessano (●); Castiglione (●); Gaglioppo (●); Greco Nero (●); Guardavalle (●); Guarnaccia (●); Inzolia (●); Magliocco Canino (●); Malvasia di Lipari (●); Mantonico Bianco (●); Mantonico Nero (●); Mennella Nera (●); Montonico Bianco Italico (●); Nerello Mascalese (●); Parmisana (●); Pecorello (□); Petrera (●); Prunesta (●); Toccarino (●); unknown (□); Uva Ruggia (●); and the reference cultivars Pinot Noir and Sangiovese (●).

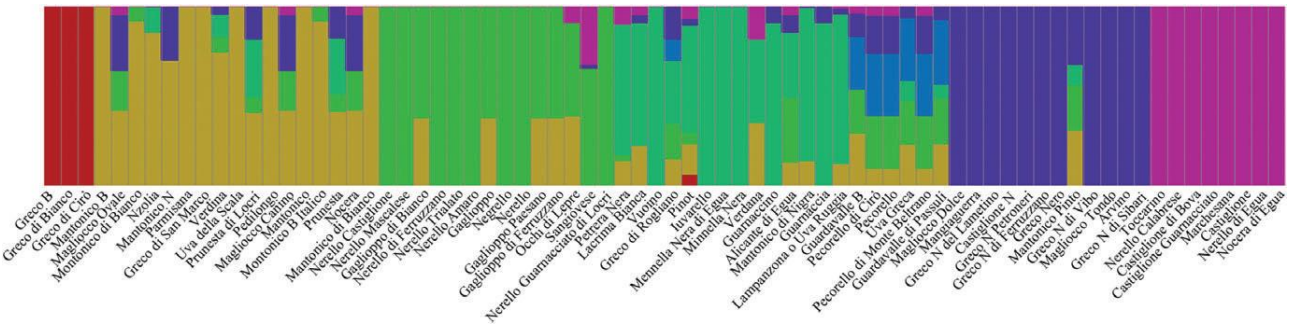

Figure 3. Admixture proportions of analysed grapevine cultivars estimated by fastSTRUCTURE (K=7). Each vertical bar represents a sample (72 genotypes belonging to the main Calabrian cultivars). The colour proportion for each bar represents the posterior probability of assignment of each individual to one of six groups of genetic similarity. The range of assignment probability varies from 0 to 100%.

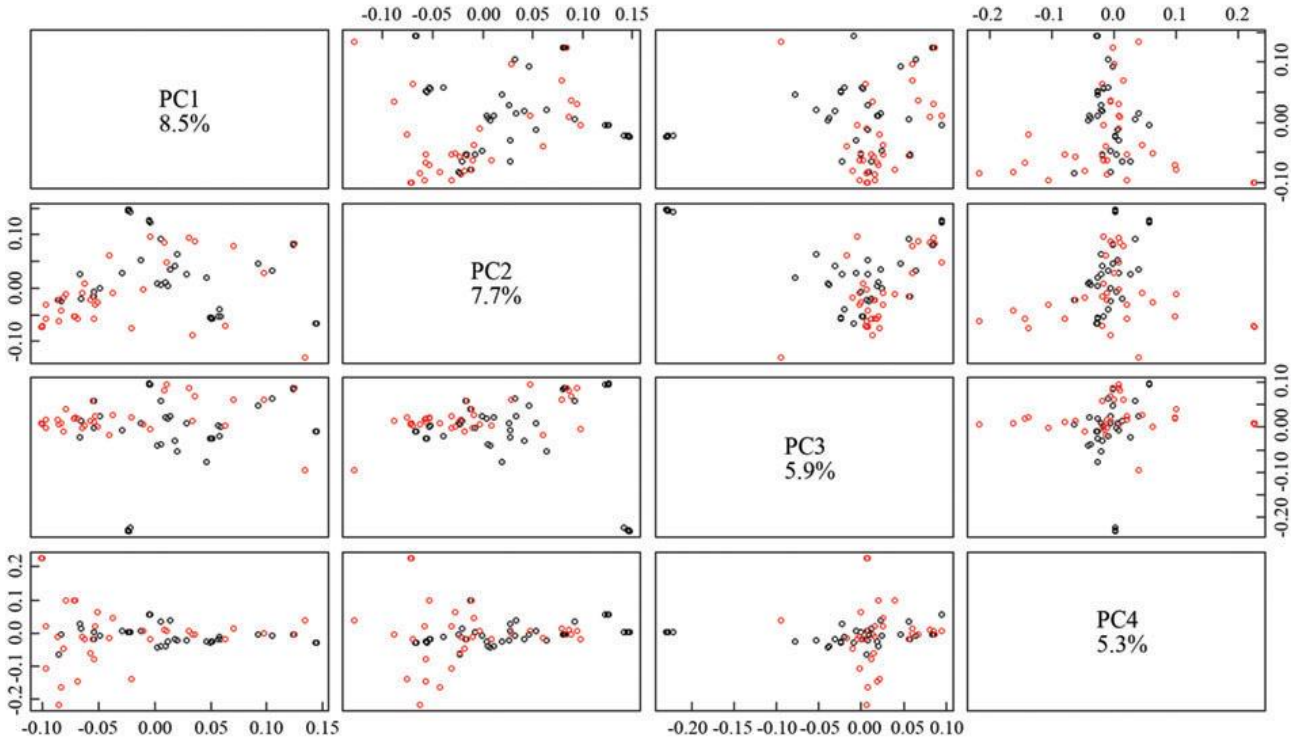

Figure 4. Principal component analysis (PCA) of the Calabrian cultivars (○) and the main Sicilian cultivars (○) previously investigated (Mercati et al. 2016). In the panel, the first four principal components (PC 1–4) are shown.

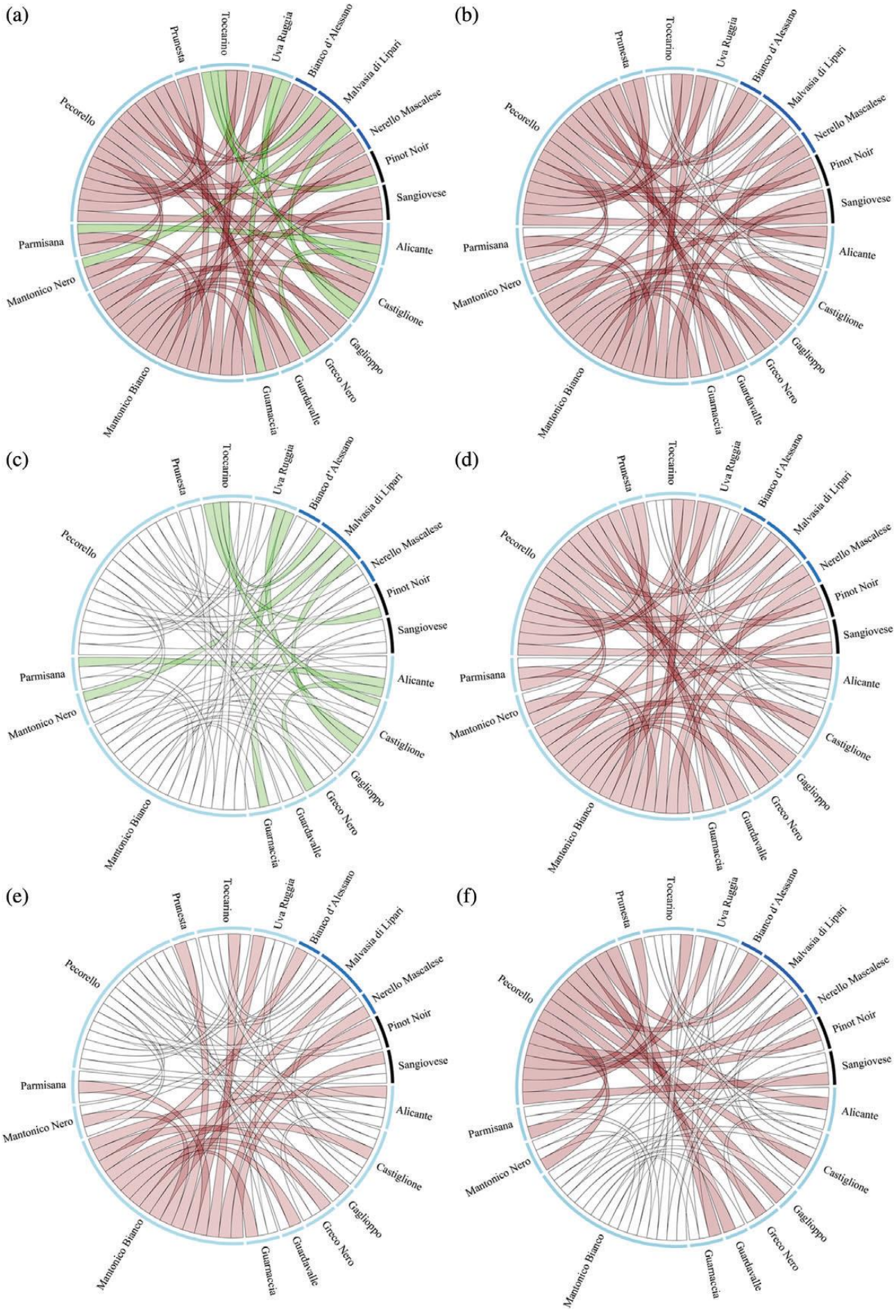
(a)
Prunesta
Toccarino
Uva Ruggia
Bianco d'Alessano
Malvasia di Lipari
Nerello Mascalese
Pinot Noir
Sangiovese
Alicante
Castiglione
Gaglioppo
Greco Nero
Guardavalle
Guarnaccia
Mantonico Bianco
Mantonico Nero
Parmisana
Pecorello
(b)
Prunesta
Toccarino
Uva Ruggia
Bianco d'Alessano
Malvasia di Lipari
Nerello Mascalese
Pinot Noir
Sangiovese
Alicante
Castiglione
Gaglioppo
Greco Nero
Guardavalle
Guarnaccia
Mantonico Bianco
Mantonico Nero
Parmisana
Pecorello
(c)
Prunesta
Toccarino
Uva Ruggia
Bianco d'Alessano
Malvasia di Lipari
Nerello Mascalese
Pinot Noir
Sangiovese
Alicante
Castiglione
Gaglioppo
Greco Nero
Guardavalle
Guarnaccia
Mantonico Bianco
Mantonico Nero
Parmisana
Pecorello
(d)
Prunesta
Toccarino
Uva Ruggia
Bianco d'Alessano
Malvasia di Lipari
Nerello Mascalese
Pinot Noir
Sangiovese
Alicante
Castiglione
Gaglioppo
Greco Nero
Guardavalle
Guarnaccia
Mantonico Bianco
Mantonico Nero
Parmisana
Pecorello
(e)
Prunesta
Toccarino
Uva Ruggia
Bianco d'Alessano
Malvasia di Lipari
Nerello Mascalese
Pinot Noir
Sangiovese
Alicante
Castiglione
Gaglioppo
Greco Nero
Guardavalle
Guarnaccia
Mantonico Bianco
Mantonico Nero
Parmisana
Pecorello
(f)
Prunesta
Toccarino
Uva Ruggia
Bianco d'Alessano
Malvasia di Lipari
Nerello Mascalese
Pinot Noir
Sangiovese
Alicante
Castiglione
Gaglioppo
Greco Nero
Guardavalle
Guarnaccia
Mantonico Bianco
Mantonico Nero
Parmisana
Pecorello

Figure 5. Circular representation of identified relationships of cultivars genotyped by the Vitis18kSNP array. The samples are arranged based on their origin: Calabrian cultivars (___); cultivars belonging to both Calabria and other Italian regions (___); reference cultivars Pinot Noir and Sangiovese (___). In the figure, parent–offspring (PO) (are shown as brown links, while first-cousin (FC), half-sibling (HS) and clones (CL) are green links, highlighting: (a) all identified relationships; (b) only PO; (c) FC, HS and CL; (d) Mantonico Bianco and Pecorello relationships; (e) Mantonico Bianco and (f) Pecorello relationships, separately..

Supplementary Information

Table S1. List of samples analysed by the Vitis18kSNP array.

Table S2. Thirty-six ampelographic Organization Internationale de la Vigne et du Vin (OIV) descriptors recorded during the 2013 and 2014 seasons for 21 out of 72 analysed accessions belonging to the main cultivars from Calabria. The expression of each trait has been recorded on 1–9 scale. For a detailed description of each trait and their expression, see the file http://www.eu-vitis.de/docs/descriptors/mcpd/WP2-DESCRIPTORS-v4.pdf.

Table S3. Microsatellite profiles at nine loci per each grapevine cultivar from Calabria. Pinot Noir and Sangiovese are included as reference cultivars.

Table S4. Homonymies and synonymies revealed by SSR analysis in germplasm grapevine collection from Calabria.

Table S5. Summary of polymorphic single nucleotide polymorphism (SNP) loci for each cultivar from Calabria. Only those with a sample size > 1 in the analysed panel were taken into account.

Table S6. Runs of homozygosity (ROHs) identified in the germplasm from Calabria.

Table S7. The proportion of runs of homozygosity (ROHs) in the genome of studied cultivars. Only 14 out of 21 cultivars showing ROHs are listed.

Table S8. Overlapping regions in the analysed germplasm collection.

Table S9. Parentage analysis and relationship category assignment for the main grapevine cultivars from Calabria and two international-national reference cultivars, Pinot Noir and Sangiovese, obtained by allelic profiles of 10 041 single nucleotide polymorphisms (SNPs). In bold: reference cultivar.

Table S10. Parentage analysis and relationship category assignment for the main grapevine cultivars from Calabria and two international-national reference cultivars, Pinot Noir and Sangiovese, obtained by allelic profiles of 10 041 single nucleotide polymorphisms.

Figure S1. Proportion of runs of homozygosity (ROH) length category 1–3 Mb (■), 3–6 Mb (■), 6–9 Mb (■) and >12 Mb (■) identified in the main Calabrian cultivars. Based on published *Vitis vinifera* genetic map [genome size 504.6 Mb, 19 linkage groups covering 1276 cM (Velasco et al. 2007)], all ROH categories identified are > 1*cM*.

Figure S2. Biplots and correlation indices for the (a) neighbourhood joining (NJ) and (b) the unweighted pair group method with arithmetic mean (UPGMA) methods used to describe genetic diversity in the analysed germplasm.