


RESEARCH ARTICLE

Open Access



# SNP genotyping elucidates the genetic diversity of *Magna Graecia* grapevine germplasm and its historical origin and dissemination

Gabriella De Lorenzis<sup>1</sup>, Francesco Mercati<sup>2</sup>, Carlo Bergamini<sup>3</sup>, Maria Francesca Cardone<sup>3</sup>, Antonio Lupini<sup>4</sup>, Antonio Mauceri<sup>4</sup>, Angelo Raffaele Caputo<sup>3</sup>, Loredana Abbate<sup>2</sup>, Maria Gabriella Barbagallo<sup>5</sup>, Donato Antonacci<sup>3</sup>, Francesco Sunseri<sup>4\*</sup> and Lucio Brancadoro<sup>1\*</sup> 

## Abstract

**Background:** *Magna Graecia* is the ancient name for the modern geopolitical region of South Italy extensively populated by Greek colonizers, shown by archeological and historical evidence to be the oldest wine growing region of Italy, crucial for the spread of specialized viticulture around Mediterranean shores. Here, the genetic diversity of *Magna Graecia* grape germplasm was assessed and its role in grapevine propagation around the Mediterranean basin was underlined.

**Results:** A large collection of grapevines from *Magna Graecia* was compared with germplasm from Georgia to the Iberian Peninsula using the 18 K SNP array. A high level of genetic diversity of the analyzed germplasm was determined; clustering, structure analysis and DAPC (Discriminant Analysis of Principal Components) highlighted the genetic relationships among genotypes from South Italy and the Eastern Mediterranean (Greece). Gene flow from east (Georgia) to west (Iberian Peninsula) was identified throughout the large number of detected admixed samples. Pedigree analysis showed a complex and well-structured network of first degree relationships, where the cultivars from *Magna Graecia* were mainly involved.

**Conclusions:** This study provided evidence that *Magna Graecia* germplasm was shaped by historical events that occurred in the area due to the robust link between South Italian and Greek genotypes, as well as, by the availability of different thermal resources for cultivars growing in such different winegrowing areas. The uniqueness of this ampelographic platform was mainly an outcome of complex natural or human-driven crosses involving elite cultivars.

**Keywords:** SNP, Molecular markers, Genetic diversity, Secondary center of domestication, Parentage

## Background

Grapevine (*Vitis vinifera* L. *subsp. sativa*) is one of the most important economic fruit species in the modern world; of West Asiatic origin, it is cultivated in a wide area from Trans-Caucasus to Western Europe and around the Mediterranean Basin [1]. The wild form

(*subsp. sylvestris*) is suggested to have first appeared about 65 million years ago and its domestication was closely related to winemaking [2–4]. This process occurred about 8000 years ago and took place in the Caucasus, in an area located between the eastern coast of the Black Sea and the southern coast of the Caspian Sea [2, 5, 6]. From there, domesticated grapevines spread to south-eastern regions of the Mediterranean. During the second half of 5rd millennium BC grapevine appeared in Southern Greece and then moved to the southern Balkans, Central and Western Europe throughout South Italy [7, 8].

\* Correspondence: francesco.sunseri@unirc.it; lucio.brancadoro@unimi.it

<sup>4</sup>Dipartimento AGRARIA, località Feo di Vito snc, 89121 Reggio Calabria, Italy

<sup>1</sup>Dipartimento di Scienze Agrarie ed Ambientali, via Celoria 2, 20133 Milan, Italy

Full list of author information is available at the end of the article



Among European countries, Italy is one of the major wine producers. As stated in the OIV Statistical Report on World Vitiviniculture, the area under vines cultivation in Italy reached 690,000 ha in 2016, with 7.9 million tons of grape production for both wine and table grape markets and 50.9 million hectoliters of wine (<http://www.oiv.int/en/>). In the National Register of Italian Grapevines more than 650 varieties are included (<http://catalogoviti.politicheagricole.it/catalogo.php>), representing a slice of grapevine diversity, as assessed by genetic analysis of Italian germplasm collections, including both local varieties and the most widely known national varieties [9–15]. A high number of synonyms were detected among genotypes collected in different winegrowing areas [15–19]. However, the number of varieties included in the Italian ampelographic platform remains still large, because in the past sexual reproduction, by spontaneous crossing, was a practice as common as vegetative propagation, evident from the large number of parentage relationships identified among Italian cultivars [10, 14–16, 20–22]. The most ancient archeological evidence of viticulture in Italy dates back to the Epigravettian and Mesolithic periods, in Grotta del Romito (Cosenza, Calabria) and Grotta dell'Uzzo (Trapani, Sicily), respectively, where seeds of wild grapevines have been identified [23, 24]. Sicily, together with the other regions of southern Italy (Calabria, Campania, Basilicata and Puglia), played a key role in the introduction of viticulture to Italy during Greek colonization and its next spread along the Italian Peninsula to reach southern France (Marseilles) and western Spain [25]. Based on cultural and historical references, the first varieties introduced to southern Italy were: i) “Biblia”, imported by the Siracusa king, Pollis d'Argo, from the north-eastern Aegean; ii) “Morghio”, introduced to the southern Bruzio (now known as Calabria); iii) “Lagaria”, introduced to the Metaponto area (Basilicata) [26]. The co-existence of these with the autochthonous cultivars would help to shape the complex ampelographic platform of the Italian Peninsula, as a region of intermixing and exchange of varieties, which resulted in an admixed genetic structure [12, 27]. These authors addressed a lack of structuring in the Italian grapevine germplasm related to historical events which occurred in this country (Greek colonization, Roman Empire, Spanish colonization) over the course of centuries, probably due to the continual exchange of grape plants inside and outside the Italian Peninsula.

According to archaeobotanical data, Italian grapevine domestication was mainly determined by local grape populations, sharing a genetic pool with varieties from the Hellenic world, where viticulture had already reached a high level of specialization [28, 29]. As a consequence, a secondary domestication center of grapevine

arose, as a long-term process of hybridization and selection of suitable genotypes. This hypothesis is supported by the finding of seed remains exhibiting intermediate traits between the two subspecies *sylvestris* and *sativa*, on the islands of Filicudi and Salina (Sicily), referring to the Middle bronze age [29, 30]. The first evidence of a developed cultivation system, dating back to the Middle Bronze Age, was found at sites in Strepparo and Cento Moggie (Caserta, Campania), where grapevine branches were discovered [31].

The biodiversity of southern Italian grapevine germplasm has been widely investigated by SSR (Simple Sequence Repeat) to study genetic variability [9–11, 14, 32], identify homonymies/synonymies [10, 14, 16] and parental relationships [14–16, 19, 20]. Structure and genetic diversity of local germplasm from southern Italy matches its historical and geographical background, and many synonyms, homonyms and parent-offspring relationships have been confirmed. Moreover, the identification of cultivars, such as Sangiovese, showing several relationships with southern Italian germplasm strengthens the genetic complexity of this ampelographic platform [14–16, 20, 21, 32, 33].

Recently, programmes of sequencing and re-sequencing of the grape genome have generated a database including an extensive number of single nucleotide polymorphisms (SNP), useful for setting up different genotyping SNP-panels developed for both SNPlex™ and chip array strategies [13, 34–36]. An initial set of 10 K SNP loci obtained from 17 grape DNA samples (10 cultivated *V. vinifera* and 7 wild *Vitis* species) was developed by Myles et al. [36]. The GrapeReSeq Consortium developed the Vitis18KSNP chip array, holding 18,775 SNPs chosen from *V. vinifera* and *Vitis* spp. genotypes [35]. Finally, Marrano et al. [37] described a new set of 37 K SNP in a grapevine collection of cultivated and wild accessions through a novel protocol of restriction-site associated DNA sequencing.

The usefulness of SNP sets has been demonstrated despite the bi-allelic nature of this kind of molecular marker. They were proven adequate to investigate genetic variability, discriminating among *V. vinifera* populations, between wild and cultivated compartments of *V. vinifera* and among wild *Vitis* species [15, 33–35, 37–41]. Furthermore, SNP were able to infer genetic structure [13, 38] and to identify kinships [15, 34, 38, 42]. In addition, the advantages of SNP genotyping are: i) high reproducibility among laboratories, indeed normalization with reference varieties is not required; ii) locus availability, thousands or millions of SNP can be retrieved from genome sequences; iii) high-throughput, multiplexing hundreds or thousands of loci in one chip; iv) automatization, sample processing may be completely automated.

*Magna Graecia* is the ancient name of southern Italy (nowadays the regions of Basilicata, Calabria, Campania, Puglia and Sicilia) colonized by the Greeks during the eighth century BC. This area was influenced by Greek civilization, in terms of customs and traditions, such as language, religious rites and agriculture, including viticulture and its varieties [25]. Here, a large germplasm collection of grapevine accessions originating from *Magna Graecia* were genotyped by Vitis18kSNP chip array to identify synonymies/homonymies, to investigate genetic diversity, population structure and parentage. To highlight the east-to-west gene-flow between the primary (Caucasus region) and the secondary domestication centers, a SNP dataset with genotypes from Georgia up to the Iberian Peninsula were included in the analysis.

## Methods

### Plant material

A total of 140 grapevine accessions, originating from southern Italy, Greece and the Eastern Mediterranean Sea (Bosnia and Herzegovina, Croatia, Lebanon, Montenegro, Slovenia, Turkey) were genotyped by Vitis18kSNP chip array (Table 1; Fig. 1). The Italian samples were from five regions suited to viticulture from ancient times: Basilicata, Calabria, Campania, Puglia and Sicilia. To this set of samples, SNP profiles of 10 and 32 cultivars from Sicilia and Calabria, respectively, were included [15, 38], as well as SNP profile of Aglianico [41]. Aleatico, Moscato bianco, Pinot noir and Sangiovese profiles [38] were finally included in the dataset as reference varieties. A detailed list of plant material is reported in Additional file 1a, including passport data of accessions and SSR-molecular data at 9 loci (VrZag62, VrZag79, VVMD5, VVMD7, VVMD25, VVMD27, VVMD28, VVMD32, VVS2; [43–46]) already published or coming from our private databases. This sample set, encompassing 187 genotypes described above, is named sample set #1 from here on.

To investigate the role of *Magna Graecia* germplasm into the domestication and dissemination process of

grapevine, our sample set (#1) was added with SNP profiles already available. This second sample set (#2) included genotypes from sample set #1, SNP profile of 42 varieties coming from Georgia already described in De Lorenzis et al. [39] and a subset of 478 SNP profiles of cultivars coming from Georgia, Turkey, Cyprus, Greece, Lebanon, Balkans, Italy, France and Iberian Peninsula [22] (Additional file 1b). This second sample set, encompassing 709 genotypes described above, is named sample set #2 from here on.

### DNA extraction and genotyping

Genomic DNA was extracted from 100 mg of young leaf tissue by the DNeasy™ Plant Mini Kit (Qiagen - Hilden, Germany), following the manufacture's instructions. The quality (260/230 and 260/280 ratios) and quantity of DNA extracted were checked using the NanoDrop Spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and the Quant-iT dsDNA HS assay kit for Qubit 3.0 Fluorometer (Thermo Fisher Scientific), respectively.

The Vitis18kSNP array (Illumina Inc., San Diego, California), containing 18,071 SNPs, were used to genotype the 195 samples. The amplifications were performed on 200 ng of genomic DNA by the laboratory of Fondazione Edmund Much (San Michele all'Adige, Trento, Italy). SNP calls were scored with Genotyping Module 1.9.4 of the GenomeStudio Data Analysis V2011.1 software (Illumina Inc.). SNP loci showing call quality values (p50GC) lower than 0.54 were removed from the final dataset, as well as loci having GenTrain (GT) score values lower than 0.6. [39] and those with a percentage of missing data higher than 20% and minor allele frequency (MAF) lower than 0.05.

### Data analysis

#### Sample set #1

On the first sample set (#1), genetic distances and variability were assessed. The genetic variability was estimated by observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ) [47], the minor allele frequency (MAF), calculated by PEAS 1.0 software [48], and inbreeding coefficient (F), determined by R 3.4 software [49]. The genetic distances between genotypes was calculated by *poppr* [50] package implemented in R. The distance matrix was set up on Nei's distances [51] and Unweighted Pair Group Method with Arithmetic Mean (UPGMA) algorithm was used for clustering. The circular dendrogram was plotted by using MEGA 5.0 software [52].

#### Sample set #2

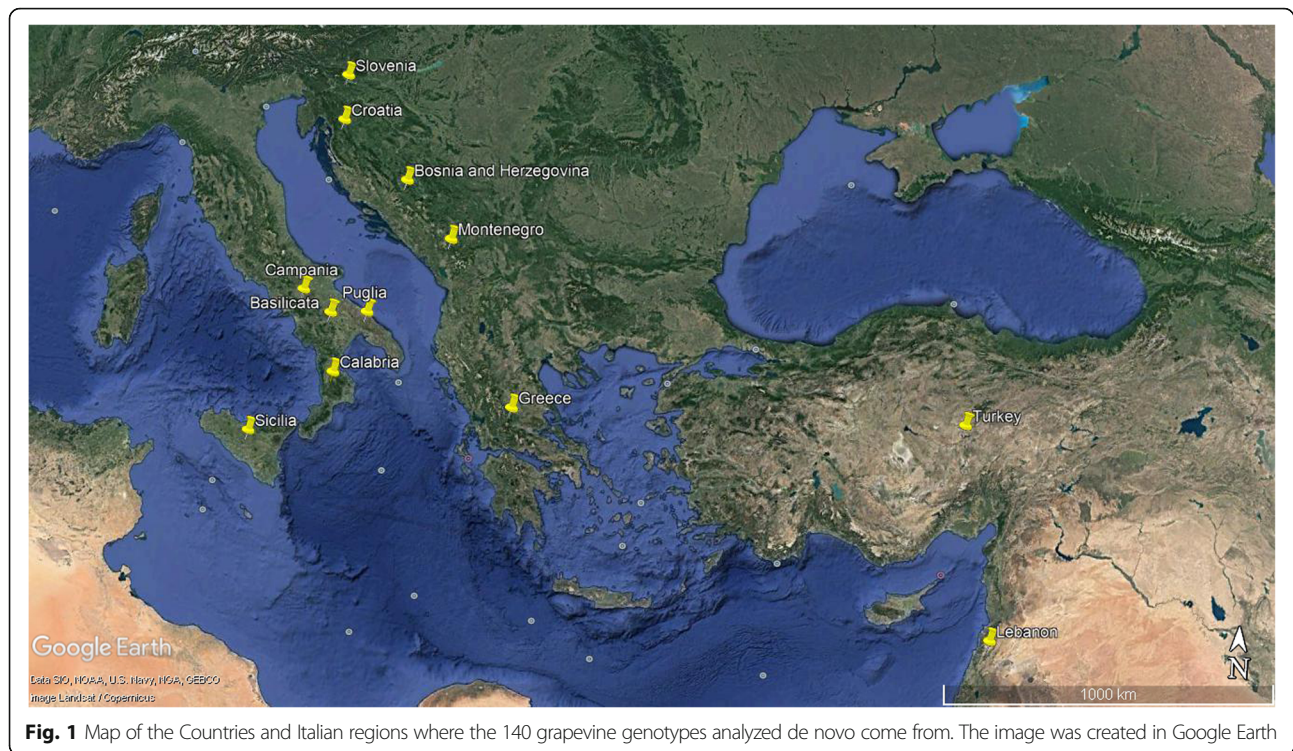
On sample set #2,  $H_e$ ,  $H_o$ , MAF and F indexes and cluster analysis were performed as described for sample set #1. The dendrogram was displayed as topology in MEGA 5.0 software [52], where the relationships among genotypes were viewed ignoring the branch lengths.

**Table 1** Number of grapevine accessions genotyped de novo by 18 K SNP array and arranged based on their geographical origin

Populations	Number of genotypes
South Italy	111
Basilicata	15
Calabria	4
Campania	42
Puglia	25
Sicilia	25
Eastern Mediterranean Sea <sup>a</sup>	29
Total	140

<sup>a</sup>Genotypes originated in Bosnia and Herzegovina (4), Croatia (1), Greece (20), Lebanon (1), Montenegro (1), Slovenia (1) and Turkey (1)





**Fig. 1** Map of the Countries and Italian regions where the 140 grapevine genotypes analyzed de novo come from. The image was created in Google Earth

The genetic structure was investigated in more detail with a Bayesian model-based clustering algorithm implemented in *tess3* R package [53]. Based on SNP profiles, individuals were assigned to  $K$  populations, estimating the membership proportion for each genotype at each  $K$ . The program was run for ancestral population numbers ranging from  $K = 1$  to  $K = 10$ . The algorithm was repeated 10 times for each  $K$  value, lambda value for the spatial regularization parameter was 1, the method chosen was “projected.ls” (an alternating projected least squares algorithm), with a maximum iterations number of the optimization algorithm up to 200 and a tolerance (value corresponding to the stopping criteria of the optimization algorithm) of  $1e-05$ . The most likely  $K$  value was estimated inspecting the cross-validation curve and the results (membership proportion for each genotype at each  $K$ ) were interpolated on a geographic map.

A Discriminant Analysis of Principal Components (DAPC; [54]) was performed to identify genetic clusters using the package *adegenet* of R software [55]. The maximum number of clusters was set to 10 and the number of axes considered in the Principal Component Analysis (PCA) was set to 60. The results were viewed as two-dimension scatter plot.  $H_e$  values were calculated and the genetic differentiation among DAPC clusters was validated by pairwise Nei’s standard genetic distance [52, 56] and pairwise  $F_{st}$  analysis [57], performed in R, using *nei.dist* function of *Poppr* package [50] and *pp.fst*

function of *HierFstat* package [58], respectively. A Mantel test for Isolation-by distance (IBD) relationship was performed using the R software package *adegenet* [55] to calculate significant correlation among matrices of genetic (Nei’s and  $F_{st}$ ) and geographical distances among the populations. Four-population test (A,B; C,D) implemented in TreeMix software [59] was used to test the gene flow among the clusters identified by DAPC. Standard errors for  $f$ -statistics were calculated in blocks of 500 SNPs (i.e.  $-K$  500). A significant non-zero Z-score indicated gene flow between A,B and C,D. Higher values showed evidence of gene flow in the tree.

In order to infer relationships among non-redundant individuals, the identity-by-descent (IBD) index was calculated by PLINK 1.07 software [60]. IBD index counts for the probability that two genotypes are descended from single ancestral genotype and not identical by chance. MAF value was set at 0.1, while  $r^2$  of linkage disequilibrium value at 0.05. The relationships among pair of genotypes were assigned taking into account four parameters:  $Z_0$  (probability of sharing 0 IBD allele identical-by-descent),  $Z_1$  (probability to share 1 IBD allele),  $Z_2$  (probability to share 2 IBD alleles), and PI-HAT [(the relatedness measure measured as  $PI-HAT = P(IBD = 2) + 0.5 \times P(IBD = 1)$ ). A first-degree relationship (parent-offspring) was assigned to those pair of genotypes showing  $Z_0$  and  $Z_2$  values similar to 0,  $Z_1$  similar to 1 and PI-HAT to 0.5, a second-degree relationship when  $Z_0$  and  $Z_2$  showed values similar to 0.25,  $Z_1$  and

PI-HAT to 0.5. An overview of first- and second-degree relationships among samples was displayed by circular visualization, obtained by Circlize R package [61]. The network of first degree relationships was obtained by R Network package [62].

## Results

### Genetic diversity of *Magna Graecia* germplasm

The genetic diversity among southern Italian grapevine accessions and those coming from Eastern Mediterranean Sea (Bosnia and Herzegovina, Croatia, Greece, Lebanon, Montenegro, Slovenia, Turkey), the so called sample set #1, were investigated by a high-throughput genotyping system based on SNP chip array, the Vitis18kSNP array. The final dataset counted 11,023 loci (Additional file 2) after the removal of: i) 73 loci (less than 1%) that did not amplify overall the genotypes; ii) 4616 loci (25%) that showed GT values lower than 0.6; iii) 1937 monomorphic loci (12% of dataset after GT filtration); iv) 495 loci (4% of the resultant dataset) with a MAF value lower than 0.05. All loci exhibited p50GC values higher than 0.54, while after inspection no genotypes with a percentage of non-calling loci higher than 20% were detected.  $H_o$  values were lower than the expected values from total germplasm within the sample set #1, ranging from 0.2873 (eastern Mediterranean Sea population) to 0.3060 (Italian population). Total  $H_e$  value was 0.3179 and the range varied from 0.3017 (Georgia) to 0.3471 (South Italy, Calabria) (Table 2). The overall value of MAF was 0.2206 (Table 2) and 1102 out of 11,023 SNP loci (about 10%) showed a MAF value lower than 0.100. The MAF values, similar among the different populations, ranged from 0.2176 (Greece) to 0.2270 (South Italy, Calabria) (Table 2). The values of F index (inbreeding coefficient) ranged from 0.1276 (Greece) to 0.1906 (South Italy, Calabria), with a value overall the population equal to 0.1496.

**Table 2** Summary of genetic variation statistics at 18 K SNP loci on sample set #1

Populations	$H_o^a$	$H_e^b$	MAF <sup>c</sup>	F <sup>d</sup>
South Italy	0.3060	0.3251	0.2246	0.1654
Basilicata	0.3066	0.3280	0.2240	0.1605
Calabria	0.3059	0.3471	0.2270	0.1906
Campania	0.3099	0.3175	0.2268	0.1564
Puglia	0.3009	0.3173	0.2207	0.1378
Sicilia	0.3068	0.3154	0.2247	0.1425
Eastern Mediterranean Sea	0.2873	0.3124	0.2125	0.1321
Greece	0.2958	0.3037	0.2176	0.1276
Total	0.3001	0.3179	0.2206	0.1496

<sup>a</sup>Observed heterozygosity; <sup>b</sup>Expected heterozygosity; <sup>c</sup>Minor allele frequency; <sup>d</sup>Inbreeding coefficient

A UPGMA dendrogram was built to investigate the genetic relationships among the genotypes from the 18 K SNP matrix data (Additional file 3). The range of similarity varied from 100 to about 83%. A total of 158 unique genotypes were detected, with the higher number of synonymies identified in two Italian regions, Calabria and Basilicata. Other synonymies were also noticed between genotypes from different Italian regions and from Italy and Eastern Mediterranean Sea (Kratosija and Primitivo) samples, as well. The synonymies identified among the *Magna Graecia* germplasm are listed in the Additional file 4. Among the Greek genotypes, Moschofilero and Mavrodaphni showed the same SNP profile. Cluster analysis was not able to define groups based on the geographic origin. Indeed, samples coming from different Italian regions were clustered together. Nevertheless, the Greek genotypes were clustered in a group including the most part of Eastern Mediterranean Sea samples. Pinot noir, Teran and some Italian genotypes clustered as the most distant genotypes, as outgroups.

### Genetic diversity of Mediterranean Basin germplasm

In order to compare the genetic relationship between *Magna Graecia* genotypes (included in the sample set #1) with that originating from neighboring winegrowing areas, from Georgia to the Iberian Peninsula, a second dataset called sample set #2 was built, including already available data from De Lorenzis et al. [39] and Laucou et al. [22]. This last set accounted for a total of 709 profiles at 7396 SNPs. Nineteen out of 709 genotypes were duplicated. The duplicated genotypes showed the same SNP profile at 99%. The inconsistencies were due to missing values.  $H_e$ ,  $H_o$ , MAF and F indexes were measured, subdividing samples based on geographic area and historical influences in, from east to west, Georgia, eastern Mediterranean Sea Countries (Turkey, Greece, Cyprus, Lebanon and Balkans), South Italy, North Italy, France and Iberian Peninsula (Spain and Portugal); the values per each population are reported in Table 3.

**Table 3** Summary of genetic variation statistics at 18 K SNP loci on sample set #2

Populations	$H_o^a$	$H_e^b$	MAF <sup>c</sup>	F <sup>d</sup>
Georgia	0.3082	0.3107	0.2157	0.1044
Eastern Mediterranean Sea <sup>e</sup>	0.2819	0.3067	0.2223	0.1184
South Italy	0.3191	0.3392	0.2330	0.1830
North Italy	0.3262	0.3406	0.2368	0.1611
France	0.3205	0.3393	0.2332	0.1715
Iberian Peninsula <sup>f</sup>	0.3111	0.3275	0.2263	0.1502
Total	0.3112	0.3273	0.2279	0.1481

<sup>a</sup>Observed heterozygosity; <sup>b</sup>Expected heterozygosity; <sup>c</sup>Minor allele frequency; <sup>d</sup>Inbreeding coefficient; <sup>e</sup>Turkey, Greece, Cyprus, Lebanon and Balkans; <sup>f</sup>Spain and Portugal

Eastern Mediterranean Sea and North Italy showed, respectively, the lowest and the highest values for both  $H_o$  (0.2819 and 0.3262) and  $H_e$  (0.3067 and 0.3406). Georgian population showed the lowest MAF value (0.2157), while North Italy population the highest (0.2368). About F index, the lowest value was detected for Georgia (0.1044) and the highest for South Italy (0.1830).

The genetic relationships among genotypes were detected from a SNP data matrix of 709 genotypes by using Nei's distances (Nei 1972) and UPGMA algorithm (Additional file 5). The genotypes showed similarity values ranging from 88 to 100%. Within samples from South Italy and compared to genotypes from other populations, novel synonymies were not detected in addition to those already described above for "Genetic diversity of *Magna Graecia* germplasm". Here three main clusters were highlighted, the largest included samples from Eastern Mediterranean Sea, South Italy and Iberian Peninsula (C1). The second included only samples from Georgia (C2), while the third cluster was mainly comprised of samples from North Italy and France (C3) (Additional file 5). Furthermore, a clear discrimination among genotypes coming from Eastern Mediterranean Sea, South Italy and Iberian Peninsula were observable in cluster C1, while 25% of Iberian Peninsula accessions were grouped in C3, together with the samples from France and North Italy.

#### Population structure of *Magna Graecia* and Mediterranean Basin germplasm

The likely number of ancestral genetic groups (K) within the germplasm from sample set #2 was investigated by *tess3* R package. The algorithm for cross-validation curve revealed  $K = 3$ . Fifty-nine out of 709 genotypes (around 8%) showed a percentage of membership  $> 80\%$  (Additional file 6). All these structured genotypes were from Georgia. The ancestry coefficients calculated by *tess3* were spatially interpolated into a map (Fig. 2). Although the high number of admixed genotypes, the spatial interpolation of ancestry proportions inferred assigned genotypes from Georgia, Eastern Mediterranean Sea Countries and South Italy to the same cluster. In the second cluster, samples belonging to North Italy and France were mainly included, while the third one is limited Iberian Peninsula samples.

DAPC identified five clusters: Cluster 1 included mainly French genotypes (78%); Cluster 2 was the most assorted cluster, grouping together mainly samples from eastern Mediterranean Sea countries (97%) and South Italy (59%), together with Georgian, French, North Italy and Iberian Peninsula genotypes; Cluster 3 comprised only Georgian samples (87%); Cluster 4, 65% of Iberian Peninsula accessions, while Cluster 5 comprised only samples coming from the South Italy (31%) (Fig. 3 and

Additional file 7). The plot of the first two principal components distinguished Clusters 3, 4 and 5, while the differences between Clusters 1 and 2 were less clear (Fig. 3). The five genetic clusters were not geographically restricted, except for Cluster 5, confirming a high level of genetic admixture. The minimum-spanning tree between populations based on the squared distances demonstrated an equidistance between samples from Cluster 2 and those included in Cluster 3, 4 and 5.  $H_e$  values of DAPC ranged from 0.2716 (Cluster 5) to 0.3073 (Cluster 1) (Table 4). The pairwise Nei's genetic distance ranged from 0.0036 (Cluster 2 vs. Cluster 4) to 0.100 (Cluster 3 vs. Cluster 4), while the  $F_{st}$  index showed values ranging from  $-0.633$  (Cluster 3 vs. Cluster 4) and  $-0.209$  (Cluster 2 vs. Cluster 4) (Additional file 8).

The geographical patterning suggested by the DAPC analysis was verified by a Mantel test to detect the isolation-by distance. In both analyses, a correlation between genetic (Nei's and  $F_{st}$ ) and geographical distance matrices was found (Additional file 9), with  $r = 0.751$  and  $-0.789$  for geographical distance matrix and Nei's and  $F_{st}$  matrices, respectively.

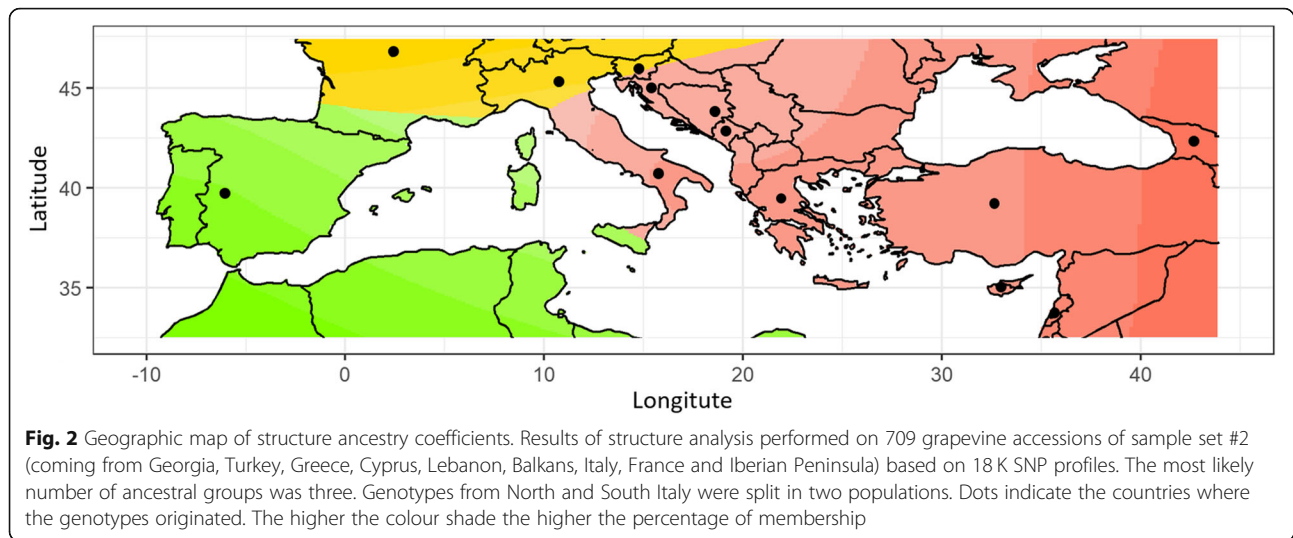
The four population test  $f_4$  (A, B; C, D) tests whether population A and B vs. population D and C can be considered distinct clades in a population tree. The groups showing the highest non-zero Z-score values, suggesting gene flow, were Cluster 1 and 5 vs. Cluster 2 and 3 (Z-score =  $-7.2797$ ) and Cluster 1 and 4 vs. Cluster 2 and 3 (Z-score =  $-9.2147$ ) (Additional file 10).

#### Parentage analysis of *Magna Graecia* germplasm

The first (PO, parent-offspring) and second-degree (grandparent-grandoffspring, half-siblings or uncle-nephew) relationships were detected among 709 genotypes included in sample set #2 to define a proposed pedigree of southern Italian samples. Z0, Z1, Z2 and PI-HAT values for PO and second-degree relationships detected among Southern Italian samples were reported in Additional file 11. Up to 82 genotypes from South Italy shared almost one PO relationship, while 100 PO relationships were detected (Additional file 11, Additional file 12). Most relationships were identified between Southern Italian genotypes. Inter-population relationships (between South Italy genotypes and other populations) were also identified, such as between Puglia and Eastern Mediterranean Sea samples (Primitivo and Blatina, Pampanuto and Plavina), Calabria and Eastern Mediterranean Sea samples (Lacrima bianca and Ladikino), Campania and Iberian Peninsula (Gloria and Breval negro).

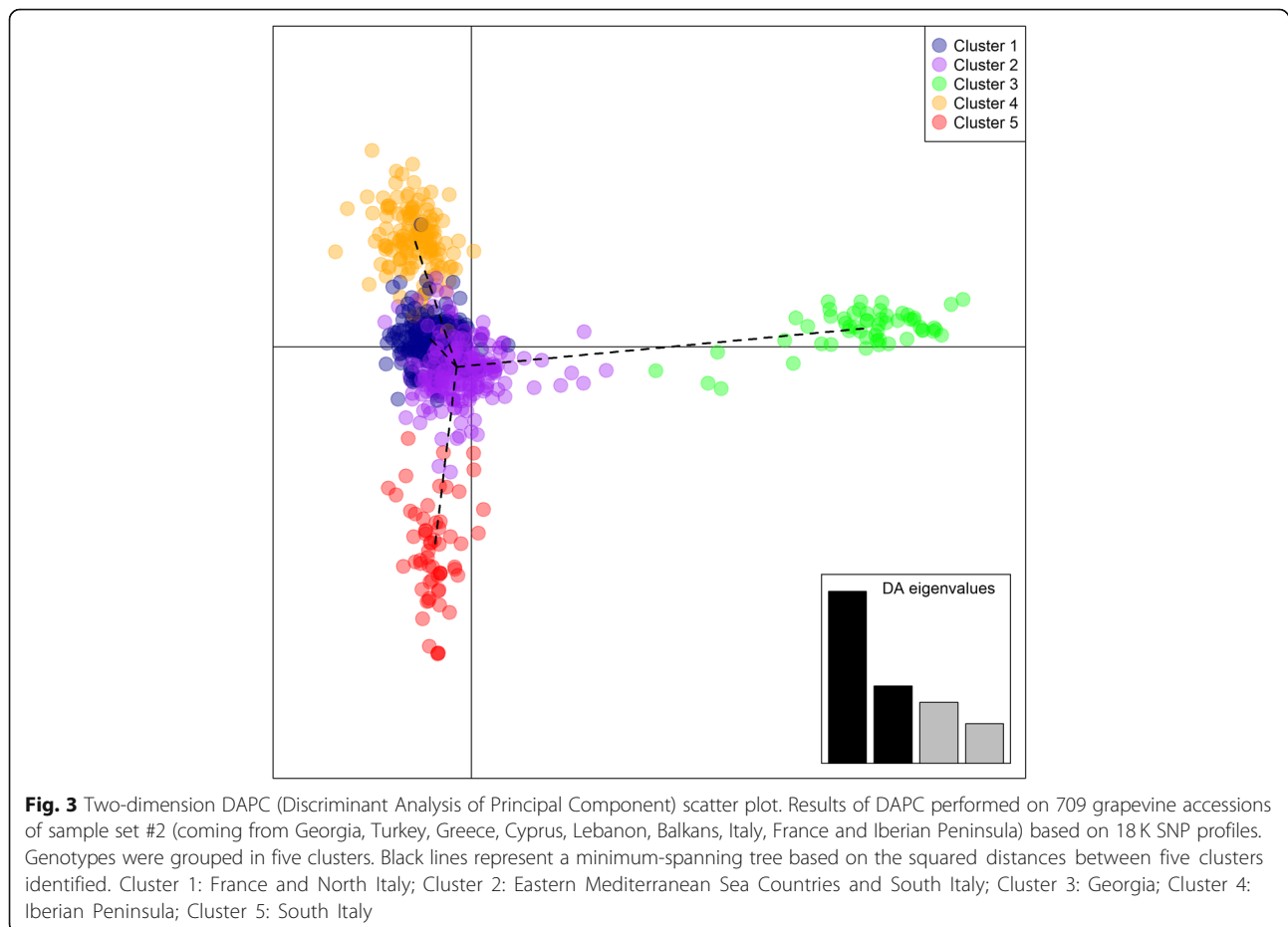
The cultivars with the highest number of relationships were Sangiovese (13) and Mantónico Bianco (10) (Fig. 4, Additional file 11, Additional file 12). Sangiovese shared kinship relations with cultivars from four out of five Italian regions investigated:





Basilicata (Stampacavallo, Strinto porcino), Calabria (Castiglione di Bova, Gaglioppo, Toccarino), Puglia (Negrodolce n.) and Sicily (Arbanello, Frappato, Lievuso, Lucignola, Orisi, Perricone). Further, Sangiovese showed PO relationships with Ciliegiolo [22]. Mantonico

Bianco showed first degree relationships with genotypes from Basilicata (Trebbiano antico), Calabria (Gaglioppo, Gallico n., Guardavalle B, Occhi di Lepre, Mantonico N) and Sicily (Catarratto, Lievuso, Nerello mascalese and Nero d'Avola).



**Table 4** Expected heterozygosity ( $H_e$ ) values of five clusters inferred by DAPC

DAPC clusters	$H_e$
Cluster 1	0.3073
Cluster 2	0.2978
Cluster 3	0.3041
Cluster 4	0.2772
Cluster 5	0.2716

Cluster 1: France and North Italy; Cluster 2: Eastern Mediterranean Sea countries and South Italy; Cluster 3: Georgia; Cluster 4: Iberian Peninsula; Cluster 5: South Italy

Other genotypes showing a well-defined pedigree-network were Bombino bianco (7 PO relationships) and Aglianico (6). Bombino bianco exhibited Z0, Z1, Z2 and PI-HAT values similar to PO relationship with Bombino n., Cococciola, Colatamurro, Impigno b., Montepulciano, Sanguinella b. and Uva di Troia (Additional file 11, Additional file 12). Aglianico showed PO relationships with Aglianico antico, Aglianicone, Aglianicone di Cardile, Cannamelo n., Rossa di Moico and Rossa di San Nicola (Additional file 11, Additional file 12).

Eight genotypes showed a second degree relationship, for a total of six relationships (Additional file 11, Additional file 12). These second degree relationships have been identified among Italian samples, and between Nerello cappuccio (Sicily) and Morate from Spain.

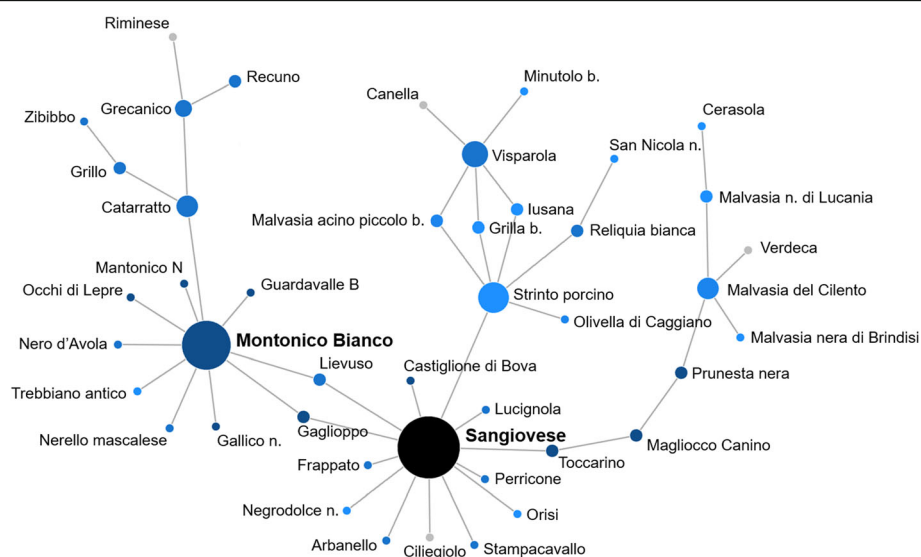
## Discussion

### *Magna Graecia* germplasm has a high genetic diversity

The 18,071 SNPs held in the Vitis18kSNP array were used to investigate the genetic diversity of grapevine

germplasm native to *Magna Graecia*. Due to its centrality in the Mediterranean Basin and political and economic contacts with other Mediterranean countries, the genetic relationships between *Magna Graecia* germplasm and those from the neighboring countries and the primary grape domestication center [2] were assessed. To accomplish genetic characterization, two different datasets were built. The first one included samples from *Magna Graecia*, Greece and some samples from the Balkans, called sample set #1 (Table 1 and Additional file 1a). The sample set #2 was built adding to the first SNP profiles of grapevines from Mediterranean area (from Georgia to Portugal) previously reported [22, 39] (Additional file 1b).

In the sample set #1, all the accessions were successfully genotyped, resulting 11,023 polymorphic SNPs (Additional file 2), after filtering loci that did not amplify overall the entire dataset, that did not reach the GT threshold value of 0.6, monomorphic and with a MAF value lower than 0.05. Even though the dataset was as-sorted, the number of monomorphic loci was quite high, reaching 12%. However, the number of loci that did not show amplification (1%) was reasonable taking into account that 25% of SNP loci included in the array were from *Vitis* genomes but not from *V. vinifera* [35]. This result is comparable with those previously reported [15, 38, 39].  $H_o$  values (of approximately 0.3000) were similar to  $H_e$  values (0.3179 across populations), suggesting that the populations can be affected by frequent in-breeding events (Table 2). These data supported previous observations based on both different [13, 33, 34, 37] and identical sets of SNPs [15, 38, 39]. The Italian population



**Fig. 4** Network of first-degree relationships of Mantonico Bianco and Sangiovese. Vertices were colored based on the geographical origin of genotypes (Italy: blue scale based on sub-populations; genotypes from Laucou et al. [22]: grey; Reference: black) and their size was scaled based on the number of first-degree relationships of each genotype



and its subpopulations showed the highest  $H_e$  values, indicating a relatively high level of genetic variability. The distribution of MAF values (Table 2) and the percentage of loci showing MAF values lower than 0.05 (4%) was quite similar to those observed using different SNP panels [13, 34, 37]. The low  $F$  values (0.1496 across populations; Table 2) were justified by the high level of heterozygosity observed, as previously stated [3, 63].

After merging genotypes from Georgia and Mediterranean Basin Countries [22, 39] with sample set #1, the resulting sample set #2 counted 709 samples genotyped at 7396 SNP loci. Duplicated genotypes showed 99% of homology in SNP profile, demonstrating the high reproducibility of this genotyping tool [38], and the ease of merging large datasets from different scientific reports [22]. Likewise, sample set #2 showed high values of heterozygosity (total  $H_o = 0.3112$ , total  $H_e = 0.3273$ ) and  $H_o$  values higher than  $H_e$  values (Table 3). The eastern Mediterranean Sea population showed the lowest heterozygosity values for both datasets (Table 2 and Table 3), while the highest genetic diversity was detected for genotypes coming from North Italy (Table 3). MAF and  $F$  values were similar between the two datasets, Georgia showed the lowest values for both MAF and  $F$  indexes, while northern and southern Italian populations showed the highest MAF and  $F$  values, respectively (Table 3). The similarity of genetic variation statistics between the two datasets suggests that those parameters are not affected by population size and number of loci.

#### The common genetic background of *Magna Graecia* germplasm

Cluster analysis on sample set #1 identified 22 cases of synonymies (Additional file 3, Additional file 4), most of which were detected among Italian samples. Some cultivars were identified in more than one region, such as Nerello mascalese (from Sicilia), also identified in Calabria and Campania. Zibibbo sampled in Campania and Sicilia, Primitivo in Campania and Montenegro (Kratosija), as already reported in other studies [10, 14, 15, 21, 64–67]. Many known synonymies were confirmed, while the remaining eight cases of synonymies were identified here for the first time (Additional file 4). The Moschofilero-Mavrodaphni synonymy was a clear misnaming, since they are supposedly different genotypes [68].

The clustering did not reflect the geographical origin of samples from South Italy, proving the common genetic background; even though some cultivars were clustered as the outgroup (Additional file 3). Moreover, the grouping of Italian and Greek samples together in the same clusters are in agreement with the historical events joining these two areas [25]. These results supported the hypothesis that during Greek colonization, Calabria and

Sicilia played an important role for evaluating the potential of varieties coming from the Eastern Mediterranean Sea and their spreading firstly in South Italy and afterwards in Etruscan Italy (Central Italy) and France [25, 69]. During the assessment, the imported varieties could be crossed with both wild and domesticated local grapevines, overlapping with a domestication process of wild grapevines already under way at that time [70].

#### The uniqueness of Georgian grapevine germplasm

The Caucasus is the place where grapevine domestication took place [71]. The genetic diversity of wild and cultivated grapevine germplasm from Georgia was extensively investigated by SSR and SNP molecular markers [14, 22, 27, 34, 72, 73]. The results demonstrated the uniqueness and originality of this germplasm due to distinctive traits compared to the other European samples, as well as the high level of heterozygosity. This uniqueness was confirmed by most of the statistics applied here. Cluster analysis and DAPC strongly separated Georgian samples from the others (Additional file 5, Fig. 3). This differentiation was also stated by pairwise Nei's and  $F_{st}$  genetic distance values (Additional file 8). Indeed, pairwise analyses including Georgian genotypes showed the highest and lowest values for Nei's and  $F_{st}$  genetic distance, respectively. Despite the high level of admixture, structure analysis identified three ancestral populations (Fig. 2). Georgian germplasm was grouped with the genotypes from the eastern Mediterranean Sea and South Italy, but appeared as unique population with well-structured genotypes, with a percentage of membership > 80% (Additional file 6). The same result was also highlighted by the spatial interpolation of the ancestry values (Fig. 2).

#### The grapevine distribution route from Georgia to Iberian Peninsula through *Magna Graecia*

Domesticated grapevine spread gradually from the Caucasus westwards via Anatolia and Greece by different Peoples and from this area into the South Italy by the Greeks [71]. In Greece, the most ancient evidence of viticulture dates back to the 5th millennium BC [74, 75], while in Italy, the beginnings of grapevine cultivation would have started during the ninth century BC [76].

Myles et al. [34] suggested the hypothesis of an east-to-west flow of genetic resources from the primary domestication center. Here, this hypothesis was strongly confirmed by the spatial interpolated ancestry coefficient map, DAPC and cluster analysis (Fig. 2, Fig. 3, Additional file 5). Georgian germplasm is distinguishable from the others, appearing closely related mainly to eastern Mediterranean Sea and South Italian populations. Gene flow was also highlighted by four population test, indicating gene flow between French/North Italy and Iberian

Peninsula genotypes (A,B; Cluster 1, Cluster 4) and South Italy/Eastern Mediterranean sea and Georgia genotypes (C,D; Cluster 2, Cluster 3) (Additional file 10). This result was confirmed by the high level of admixture detected by structure analysis (Additional file 6) and low *F<sub>st</sub>* values (Additional file 8), indicating low population structuring and limited barriers to gene flow.

Grapevines were already found to have spread from *Magna Graecia* to Spain and France [71, 77]. Nevertheless, a relationship between South Italy and the Iberian Peninsula germplasm was only detected in cluster analysis (Additional file 5), suggesting a different genetic origin of this last population. Arroyo-Garcia et al. [78] suggested a secondary domestication center in Western Mediterranean regions (Iberian Peninsula, Central Europe and Northern Africa) based on chloroplastic DNA haplotype differences. This statement was partially confirmed by our analyses. Indeed, Iberian Peninsula germplasm was grouped aside from the French germplasm by structure analysis, DAPC and clustering (Fig. 2 and Fig. 3, Additional file 5). These results can be interpreted as the existence of different genetic backgrounds contributing to the genetic make-up of current grapevines of Iberian Peninsula and France. This different genetic background may possibly be related to different grapevine *sylvestris* populations. Wild populations of different genetic background were identified in Spain, resulting from disconnected glacial refugees in the Iberian Peninsula during the last Pleistocene glaciations [79]. Individuals from these wild populations contributed to the development of germplasm from Iberian Peninsula [73]. Conversely, Central European germplasm was reported to be affected by the admixture confluence of migration routes radiating from separate refugees during the postglacial era [80].

#### Biodiversity of northern and southern Italian germplasm as a mirror of historical events

Structure population analysis applied to large grapevine collections revealed genetic groups defined mainly by subspecies (*sativa* or *sylvestris*), geographic origin and usage (wine or table grapes; [13, 27]). In this study, the main subdivision based on the geographical origin of samples was not totally captured. Focusing on whole Italian germplasm, an interesting clear discrimination among North and South Italy samples was observed. Structure, DAPC and cluster analysis grouped aside the genotypes from North and South Italy (Fig. 2, Fig. 3 and Additional file 5). Northern Italian genotypes mainly clustered with French germplasm while those from South Italy with Eastern Mediterranean Sea (clustering and DAPC) and Eastern Mediterranean Sea and Georgia (structure analysis).

In attempting to back-track the origin of this differentiation, this result can be explained by the different historical and demographical events occurred in the two areas. Italy was fragmented country until 1861 when the *Regno d'Italia* (Kingdom of Italy) was established, an united kingdom encompassing the entire Italian Peninsula. Apart the periods when the Italian territory was politically unified, such as under the domination of the Roman Empire or the Ostrogoths, different peoples have alternately occupied Italian territory over the course of the centuries. During classical age, Phoenicians and Greeks established settlements in the South Italy and the Celts inhabited the North Italy, while in the 8-9th century the Frankish Empire and Normans defeated North and South Italy, respectively, proclaiming the end of Italian political unity for the next 1300 years. Thus, rather than a division by geographic area, our data suggested a division by historical events.

Another aspect that might have been influenced the differentiation between northern and southern Italian germplasm is the different availability of thermal resources for cultivars growing in such different environments, characterized by different latitudes. Indeed, the thermal resources observed in the South Italy are similar to South-East Europe and the these observed in the North Italy are similar to North-East Europe [81].

#### Secondary grapevine domestication centers

In invasive species, genetic diversity is expected to decrease from the origin of introduction to newly invaded areas [82]. *V. vinifera* subsp. *sativa* can be considered an invasive species, though the grapevine dissemination was mainly mediated by human migrations, taking into account the first center of domestication as the origin of introduction.

From east to west, *H<sub>e</sub>* values decreased from Cluster 1 (the origin of introduction; *H<sub>e</sub>* = 0.3041) to Cluster 2 (the first newly invading area; *H<sub>e</sub>* = 0.2978). *H<sub>e</sub>* increased to a maximum for Cluster 1 (0.3073), including samples from France and North Italy (Table 4). France and North Italy showed the highest *H<sub>e</sub>* values when samples were arranged based on geographic origin (Table 3).

Where *H<sub>e</sub>* values increase along the newly invaded areas, an event that maximized genetic diversity can be hypothesized. Domestication of a species from its wild relatives is an event that maximizes the genetic diversity of such species. Secondary domestication centers along the grapevine dissemination routes were proposed. These centers refer to places where spontaneous grapevines were of interest for local people [4] and are located in the southern Balkans and Aegean Region, South Italy [70], the western Mediterranean [78], Provence-Northern Italy and Central Europe. Nevertheless, this evidence is not enough to explain the high *H<sub>e</sub>* values for French-Northern Italian genotypes.

Another aspect influencing the genetic diversity of a germplasm is the convergence of different populations in the hybrid zones. The hybrid zones are meeting areas of two populations as they expanded their ranges from separate glacial refuges. In Europe, three well-known hybrid zones were identified, in Western Central Europe along the French-German border, in the Alps and in Scandinavia [83]. Thus, the Central Europe is a melting pot, an admixture confluence of migration routes radiating from separate refuges. Postglacial range expansion, domestication of local wild individuals, conjunction of endemic varieties with highly diverse traits from distinct geographical regions migrating along the Balkan route and from the ancient Roman province of Pannonia [84], as well as socio-political events can help us to interpret the convergence of french-northern Italian grapevine germplasm.

#### The complex pedigree of *Magna Graecia* germplasm

The inherited grapevine germplasm arose by spontaneous or man-made crosses among the cultivars and their vegetative propagation across the centuries from the initial domestication to the present day. Thus, cultivars are strictly related by parentage relationships each other resulting into a complex pedigree, as already described in previous reports [12, 34, 66]. The same complex pedigree was observed in our dataset, with 82 genotypes showing a parentage relationship within the dataset (Additional file 11, Additional file 12), confirming 35 previously described PO relationships and suggesting 65 new ones. These genotypes showed such a complex number of inter-relationships that DAPC grouped them as a separate group (Cluster 5) from the remaining genotypes of Southern Italy that clustered together with the Greek samples (Cluster 2) (Fig. 3).

Most relationships were detected among South Italy cultivars, supporting the hypothesis of an intense regional exchange of plant material and a complex natural or anthropogenic breeding. Nevertheless, relationships between genotypes from South Italy and those from eastern and western countries were also detected.

The cultivars showing the highest PO relationships were Sangiovese, Mantonico Bianco, Bombino bianco and Aglianico. The genetic diversity of grape was dramatically threatened when at the end of nineteenth century phylloxera (*Daktulosphaira vitifoliae*) reached Europe from North America and devastated European vineyards. The introduction of North America non-*vinifera* cultivars as rootstock safeguarded the future of European *V. vinifera*. At the same time, together with the ever more increasing practice of non-*vinifera* rootstocks the selection and vegetative propagation of elite cultivars occurred, resulting in the further reduction of grapevine genetic diversity [3]. At the Italian scale, this

was confirmed by the identification of a high number of synonyms for the most important Italian cultivars [10–12]. Moreover, our results revealed that as the vegetative propagation of elite cultivars occurred so did the use of elite cultivars in grape breeding programs. These cultivars have shaped the grapevine ampelographic platform of the Southern Italy. The elite cultivars (such as Sangiovese and Aglianico) were varieties widespread in many important winegrowing areas or minor varieties but with large local interest (such as Mantonico Bianco and Bombino bianco). Similar results emerged when different progenies of Pinot (a worldwide cultivar) and Gouais blanc (a neglected cultivar) were identified [85].

The interest in looking for the origin and kingroup of Sangiovese, one of the most important Italian cultivars, is notable. Indeed, different pedigrees have been proposed [20, 86], as well as a large number of progenies and synonymies [12, 14–16, 32, 38]. The present study adds two new cultivars to the already long list of Sangiovese progenies. Toccarino and Strinto porcino, local cultivars from Calabria and Basilicata, respectively, attested and corroborated the strong relationship of Sangiovese with the southern Italian germplasm [12, 14–16, 20, 32, 38]. This result is strengthened by the clustering of Sangiovese within southern Italian samples (Additional file 5). Mantonico Bianco is less well known than Sangiovese, it is an ancient autochthonous cultivar from Calabria, mainly cultivated in the Locride area (nowadays the Province of Reggio Calabria) located on the Ionian Coast, one of the first provinces of *Magna Graecia*. The first evidence of Mantonico Bianco cultivation is dated back to 1600 [87]. Despite its limited spread, Mantonico Bianco showed PO relationships with cultivars from Calabria, Sicilia and Basilicata, already described in elsewhere [12, 14, 15, 19, 21, 66] and new cultivars (Gallico n., Occhi di Lepre, Nero d'Avola and Trebbiano antico).

Bombino bianco is a cultivar widespread in the South Italy, mainly in Puglia, Basilicata, Abruzzo, Lazio and Marche. It is the progenitor (together with Uva rosa antica) of Uva di Troia, Bombino nero and Impigno [88]. In the present study, four other siblings of Bombino bianco were identified: Cococciola, Colatamurro, Montepulciano and Sanguinella b.

Aglianico is an ancient grapevine cultivar mainly cultivated in Campania (biotypes Taburno and Taurasi) and Basilicata (biotype Vulture), and strongly related with the establishment of Greek colonies in Campania [89]. Here, Aglianico was identified as progenitor of several minor accessions from Basilicata and Campania. The progenitors of Aglianico are still unknown, but a putative shared parent between Aglianico and Dureza (a minor and neglected variety cultivated in the Rhône Alpes area [90]) was discovered. Nevertheless, PO

relationships were not identified between Aglianico and French genotypes.

Another interesting cultivar was Gloria, sampled in Campania and showing a PO relationship with genotypes coming from Greece, Italy, France and Spain. This admixed progeny by geographic origin mirrors the shuffling and exchanges of cuttings that occurred over the centuries following migration routes in Europe and confirms the route from Greece to France and the Iberian Peninsula through South Italy.

The name Malvasia was used to identify a group of cultivars diverse for both genetic and phenotypic profiles, cultivated in many Mediterranean countries [91–93]. In the list of PO relationships detected among the sample set #2, a large number of relations including Malvasia genotypes were detected. The majority of these genotypes were not related each other, highlighting the heterogeneity of plants labelled as “Malvasia”.

## Conclusions

Archaeological and historical data suggested that grapevine primary domestication took place in the Caucasus and from there spread to South Italy *via* Greece, in at least two separate steps. In this study, an extensive genetic characterization of *Magna Graecia* grapevine germplasm, from Georgia to the Iberian Peninsula, was carried out by 18 K SNP loci. Based on genetic analysis, *Magna Graecia* germplasm showed a high level of heterozygosity and distinctive traits, such as a common genetic background and a complex pedigree. Nevertheless, a significant degree of gene flow was observed in agreement with historical and socio-economic events that occurred in the Mediterranean Basin. These results highlighted the central role of *Magna Graecia* in the spread of grapevine through western Europe, supporting the hypothesis of an intense exchange of plant material as well as a close relationship among southern Italian cultivars and genotypes from eastern and western countries. Moreover, the genetic diversity of *Magna Graecia* germplasm has been shaped by the historical events that occurred in this area and by the variability of atmospheric driving variables (such as temperature, solar radiation and humidity) that trigger selective pressures and determine productivity, quality and territorial specificity of agroforestry productions.

## Additional files

- Additional file 1:** a Passport data of 187 grapevine genotypes coming from Italy (Basilicata, Calabria, Campania, Puglia and Sicily), Eastern Europe (Bosnia and Herzegovina, Montenegro, Slovenia, Turkey) and Greece, the so-called sample set #1, and allele profiles at 9 SSR loci (when available).  
b List of genotypes added to the sample set #1 to obtain the sample set #2. The genotypes included belong to Georgia, Bosnia and Herzegovina,

Cyprus, Spain, France, Greece, Italy, Portugal and Turkey (from De Lorenzis et al. [39] and Laucou et al. [22]). (XLSX 49 kb)

**Additional file 2:** SNP profiles of 187 grapevine accessions (sample set #1), genotyped at 18 K loci. Original dataset was filtered based on SNP call quality and GenTrain score: samples with low SNP call quality ( $p50GC < 0.54$ ) were removed from the analysis and only SNPs with a GenTrain score higher than 0.6 were retained. “A”: homozygous for dominant allele; “B”: heterozygous for recessive allele; “?”: missing data. (XLSX 7910 kb)

**Additional file 3:** UPGMA dendrogram of 187 grapevine genotypes analyzed by 18 K SNP array. The samples were marked based on their geographical origin. South Italy: blue dots; Eastern Mediterranean Sea: red dots; Reference: black dots. (TIFF 1630 kb)

**Additional file 4:** List of synonymies identified among the *Magna Graecia* germplasm (sample set #1) analyzed by 18 K SNP array. (DOCX 13 kb)

**Additional file 5:** Topology of UPGMA dendrogram of 709 grapevine accessions genotyped with 18 K SNP array. Georgia: violet branch lines; Eastern Mediterranean Sea: red branch lines; South Italy: blue branch lines; North Italy: dark green branch lines; France: light green branch lines; Iberian Peninsula: cyan branch lines. C1, C2, C3: main clusters identified. (TIF 2662 kb)

**Additional file 6:** Ancestry values at  $K = 3$  inferred by structure analysis on 709 grapevine accessions (sample set #2) coming from Georgia, Eastern Mediterranean Sea Countries (Turkey, Greece, Cyprus, Lebanon, Balkans), Italy (North and South), France and Iberian Peninsula (Spain and Portugal) genotyped at 18 K loci. (XLSX 56 kb)

**Additional file 7:** List of five cluster inferred by DAPC on 709 grapevine accessions of sample set #2 genotyped at 18 K loci. (XLSX 27 kb)

**Additional file 8:** Nei’s standard genetic distance (below the diagonal) and  $F_{st}$  index (above the diagonal) calculated on five clusters inferred by DAPC. (DOCX 12 kb)

**Additional file 9:** Graphical representation of correlations between genetic distances and geographical distance. A: Nei’s genetic distance; B:  $F_{st}$  genetic distance. Red: high correlation between genetic and geographical distances; Yellow: medium correlation between genetic and geographical distances; Blue: low correlation between genetic and geographical distances. Dots represent distance values between two populations detected by DAPC. (TIFF 1753 kb)

**Additional file 10:** Summary of four-population tests on five clusters inferred by DAPC. (DOCX 13 kb)

**Additional file 11:** Parent-offspring (PO) and second degree ( $2^{\circ}$ ) relationships identified for South Italy cultivars genotyped at 18 K SNP loci. Z0: probability to share 0 allele; Z1: probability to share 1 allele; Z2: probability to share 2 alleles; PI HAT: probability to be identical by descendent. (XLSX 18 kb)

**Additional file 12:** Circular representation of first (red links) degree relationships identified for South Italy cultivars genotyped at 18 K SNP loci. The samples are arranged based on their geographic origin. Italy: blue scale (based on sub-populations); Eastern Mediterranean Sea Countries: red; genotypes from Laucou et al. [22]: grey; Reference: black. (TIFF 19774 kb)

## Abbreviations

DAPC: Discriminant Analysis of Principal Components; F: Inbreeding coefficient; GT: GenTrain; He: Expected heterozygosity; Ho: Observed heterozygosity; IBD: Identity-by-descent; MAF: Minor allele frequency; p50GC: Call quality values; PCA: Principal Component Analysis; PI-HAT: The relatedness measure measured as  $PI-HAT = P (IBD = 2) + 0.5 \times P (IBD = 1)$ ; SNP: Single nucleotide polymorphism; SSR: Simple Sequence Repeat; UPGMA: Unweighted Pair Group Method with Arithmetic Mean; Z0: Probability of sharing 0 IBD allele identical-by-descent; Z1: Probability to share 1 IBD allele; Z2: Probability to share 2 IBD alleles

## Acknowledgements

Authors wish to thank: Prof. Attilio Scienza and Prof. Osvaldo Failla (University of Milan, DISAA) for many inspiring discussions and Dr. Simon Pierce (DiSAA) for critically reviewing the manuscript. Authors thank Asus for providing hardware.



### Funding

This work was funded and supported by: Bando PSR Regione Calabria 2007/2013 Misura 214, Progetto "ENOTRIA"; National Project - AGER grant n. 2010-2104 "An Italian *Vitis* database with multidisciplinary approach, for exploitation and valorization of the regional genotypes".

### Availability of data and materials

The dataset supporting the results of this article is available as complementary material (Additional file 2). The SNP profiles are available at the Dryad repository under the unique identifier doi:<https://doi.org/10.5061/dryad.8f89tp3>.

### Authors' contributions

GDL, FM, CB, MFC, MGB, ARC, DA, FS and LB conceived the study, participated in the design of the study and discussed the results critically. GDL, FM, CB, AL, AM, LA performed DNA extraction, SSR and SNP genotyping and data analysis. GDL drafted the manuscript. GDL, FM, CB, MFC, FS and LB revised critically it. All the authors read and approved the final manuscript.

### Ethics approval and consent to participate

Not applicable

### Consent for publication

Not applicable

### Competing interests

Francesco Sunseri is Associate Editor of *BMC Plant Biology*.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Author details

<sup>1</sup>Dipartimento di Scienze Agrarie ed Ambientali, via Celoria 2, 20133 Milan, Italy. <sup>2</sup>Istituto di Bioscienze e Biorisorse CNR, Corso Calatafimi 414, 90120 Palermo, Italy. <sup>3</sup>Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria, Centro di ricerca Viticoltura ed Enologia, CREA-VE, via Casamassima 148, 70010 Turi, Bari, Italy. <sup>4</sup>Dipartimento AGRARIA, località Feo di Vito snc, 89121 Reggio Calabria, Italy. <sup>5</sup>Dipartimento di Scienze Agrarie, Alimentari e Forestali, viale delle Scienze 11, 90128 Palermo, Italy.

Received: 30 July 2018 Accepted: 27 November 2018

Published online: 06 January 2019

### References

- Arnold C, Gillet F, Gobat JM. Situation de la vigne sauvage *Vitis vinifera* subsp. *sylvestris* en Europe. *Vitis*. 1998;37:159–70.
- McGovern PE, Rudolph HM. The analytical and archaeological challenge of detecting ancient wine: two case studies from the ancient near east. In: McGovern PE, Fleming SJ, Katz SH, editors. *The origins and ancient history of wine*. New York: Gordon and Breach; 1996. p. 57–67.
- This P, Lacombe T, Thomas MR. Historical origins and genetic diversity of wine grapes. *Trends Genet*. 2006;22:511–9.
- Forni G. The origin of "Old World" viticulture. In: Maghradze D, Rustioni L, Scienza A, Turok J, Failla O, editors. *Caucasus and Northern Black Sea Region*. Quedlinburg: Julius Kuhn-Institut; 2012. p. 27–38.
- Zohary D, Hopf M. *Domestication of plants in the old world*. New York: Oxford University Press; 2000.
- McGovern P, Jalabadze M, Batiuk S, Callahan MP, Smith KE, Hall GR, Kvavadze E, Maghradze D, Rusishvili N, Bouby L, Failla O, Cola G, Mariani L, Boaretto E, Bacilieri R, This P, Wales N, Lordkipanidze D. Early Neolithic wine of Georgia in the South Caucasus. *Proc Natl Acad Sci U S A*. 2017;114:E10309–18.
- Hopf M. South and Southwest Europe. In: Van Zeist W, Wasylkova K, Behre KE, editors. *Progress in old world palaeoethnobotany*. Rotterdam-Brookfield: Balkema; 1991. p. 274–7.
- Kroll H. Südosteuropa. In: Van Zeist W, Wasylkova K, Behre KE, editors. *Progress in old world palaeoethnobotany*. Rotterdam-Brookfield: Balkema; 1991. p. 161–77.
- Costantini L, Monaco A, Vouillamoz JF, Forlani M, Grando MS. Genetic relationships among local *Vitis vinifera* cultivars from Campania. *Vitis*. 2005;44:25–34.
- Carimi F, Mercati F, Abbate L, Sunseri F. Microsatellite analyses for evaluation of genetic diversity among Sicilian grapevine cultivars. *Genet Resour Crop Evol*. 2010;57:703–19.
- Carimi F, Mercati F, De Michele R, Fiore MC, Riccardi P, Sunseri F. Intra-variety genetic diversity of the grapevine (*Vitis vinifera* L.) cultivar 'Nero d'Avola' as revealed by microsatellite markers. *Genet Resour Crop Evol*. 2011;58:967–75.
- Cipriani G, Spadotto A, Jurman I, Di Gaspero G, Crespan M, Meneghetti S, Frare E, Vignani R, Cresti M, Morgante M, Pezzotti M, Pe E, Policriti A, Testolin R. The SSR-based molecular profile of 1005 grapevine (*Vitis vinifera* L.) accessions uncovers new synonymy and parentages, and reveals a large admixture amongst varieties of different geographic origin. *Theor Appl Genet*. 2010;121:1569–85.
- Emanuelli F, Lorenzi S, Grzeskowiak L, Catalano V, Stefanini M, Troglio M, Myles S, Martinez-Zapater J, Zyprian E, Moreira FM, Grando MS. Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. *BMC Plant Biol*. 2013;13:39.
- De Lorenzis G, Las Casas G, Brancadoro L, Scienza A. Genotyping of Sicilian grapevine germplasm resources (*V. vinifera* L.) and their relationships with Sangiovese. *Sci Hortic*. 2014;169:189–98.
- Sunseri F, Lupini A, Mauceri A, De Lorenzis G, Araniti F, Brancadoro L, Dattola A, Gullo G, Zappia R, Mercati F. 2018. 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. *Aust J Grape Wine R* 2018, 24:345–359.
- Crespan M, Calò A, Giannetto S, Sparacio A, Storchi P, Costacurta A. Sangiovese' and 'Garganega' are two key varieties of the Italian grapevine assortment evolution. *Vitis*. 2008;47:97–104.
- De Mattia F, Imazio S, Grassi F, Lovicu G, Tardaguila J, Failla O, Maitti C, Scienza A, Labra M. Genetic characterization of Sardinia grapevine cultivars by SSR markers analysis. *J Int Sci Vigne Vin*. 2007;41:175–84.
- Raimondi S, Ruffa P, De Lorenzis G, Imazio S, Fiori S, Failla O, Schneider A. Detection of grapevine synonyms in Lombardy and Piedmont regions (northern Italy). *Vitis*. 2015;54:31–6.
- Crespan M, Storchi P, Migliaro D. Mantónico bianco grapevine cultivar is the second parent of the Sicilian Catarratto. *Am J Enol Vitic*. 2017. <https://doi.org/10.5344/ajev.2016.16068>.
- Bergamini C, Caputo AR, Gasparro M, Perniola R, Cardone MF, Antonacci D. Evidences for an alternative genealogy of 'Sangiovese'. *Mol Biotechnol*. 2013;53:278–88.
- Gasparro M, Caputo AR, Bergamini C, Crupi P, Cardone MF, Perniola R, Antonacci D. Sangiovese and its offspring in southern Italy. *Mol Biotechnol*. 2013;54:581–9.
- Laucou V, Launay A, Bacilieri R, Lacombe T, Adam-Blondon AF, Bérard A, Chauveau A, de Andrés MT, Hausmann L, Ibáñez J, Le Paslier MC, Maghradze D, Martinez-Zapater JM, Maul E, Ponnaiah M, Töpfer R, Péros JP, Boursiquot JM. Extended diversity analysis of cultivated grapevine *Vitis vinifera* with 10K genome-wide SNPs. *PLoS One*. 2018;13:e0192540.
- Costantini L. Semi e carboni del Mesolitico e Neolitico della Grotta dell'Uzzo, Trapani. *Quaternaria*. 1982;23:233–47.
- Cattani L, Colonese A, Ricciardi S. Ricostruzione climatica e ambientale di Grotta del Romito: analisi palinologiche e malacofaunistiche dei livelli epigravettiani C e D. *Atti della XXXVII riunione scientifica dell'Istituto Italiano di Preistoria e Protostoria della Calabria - Scalea - Papisidero - Praia a Mare, Il, Tortona, 29 settembre/4 ottobre 2002*, 637–640.
- Buono R, Vallariello G. Introduzione e diffusione della vite (*Vitis vinifera* L.) in Italia. *Delpinoa*. 2002;44:39–51.
- Forni G. La tecnologia e l'economia vitivinicola della Magna Graecia e della Sicilia antica: il contributo conoscitivo delle Tavole di Eraclea, della paleontologia linguistica dei dialetti meridionali, dell'archeologia. In: *Alle radici della civiltà del vino in Sicilia. Mito Archeologia e Storia*. Edited by Failla O, Forni G. Menfi, Cantine Settesoli 1996, 3–63.
- Bacilieri R, Lacombe T, Le Cunff L, Di Vecchi-Staraz M, Laucou V, Genna B, Péros JP, This P, Boursiquot JM. Genetic structure in cultivated grapevines is linked to geography and human selection. *BMC Plant Biol*. 2013;13:25.
- Lentini A. Archaeobotany brewing and winemaking in Mediterranean Basin and Transcaucasus area. In: Flourentzos P, Belgiojorno MR, Micosia LA, editors. *Cyprus in the Prehistory of Wine: de Strobel publisher, Nicosia; 2005*. p. 11–6.
- Martinelli MC, Fiorentino G, Prosdociami B, D'oronzio C, Levi ST, Mangano G, Stellati A, Wolff N. Nuove ricerche nell'insediamento sull'Istmo di Filo Braccio a Filicudi. Nota preliminare sugli scavi 2009. *Origini XXXII*. 2010; 4:285–314.

30. Fiorentino G, Colaianni G, Grasso AM, Stellati A. Caratteristiche del paleoambiente e modalità di sfruttamento dei vegetali a salina nel corso dell'età del bronzo. In: Martinelli MC, Messina, Rebus E, editors. Il villaggio dell'Età del Bronzo di Portella nelle Isole Eolie (Scavi 2006–2008); 2011. p. 233–40.
31. Castiglioni E, Rottoli M. Capua (Caserta). Località Strepparo e Cento Moggie. Scavi nell'area del Circa. Resti botanici da un pozzo dell'Età del Bronzo. *Bollettino di Archeologia*. 2001;1:37–8.
32. Di Vecchi Staraz M, Bandinelli R, Boselli M, This P, Boursiquot JM, Laucou V, Lacombe T, Varès D. Genetic structuring and parentage analysis for evolutionary studies in grapevine: kin group and origin of the cultivar Sangiovese revealed. *J Am Soc Hortic Sci*. 2007;132:514–24.
33. Cabezas JA, Ibáñez J, Lijavetzky D, Vélez D, Bravo G, Rodríguez V, Carreño I, Jermakow AM, Carreño J, Ruiz-García L, Thomas MR, Martínez-Zapater JM. A 48 SNP set for grapevine cultivar identification. *BMC Plant Biol*. 2011;11:153.
34. Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F, Aradhya MK, Prins B, Reynolds A, Chia JM, Ware D, Bustamante CD, Buckler ES. Genetic structure and domestication history of the grape. *Proc Natl Acad Sci U S A*. 2011;108:3530–5.
35. Le Paslier M-C, Choise N, Bacilieri R, Bounon R, Boursiquot J-M, Bras M, Brunel D, Di Gaspero G, Hausmann L, Lacombe T, Laucou V, Launay A, Martínez-Zapater JM, Morgante M, Raj PM, Ponnaiah M, Quesneville H, Scalabrini S, Torres-Perez R, Adam-Blondin A-F: The GrapeReSeq 8k *Vitis* genotyping chip. In 9th International symposium grapevine physiology and Biotechnology International Society for Hortic Sci; 2–26 April 2013; La Serena; Book of Abstract p. 23.
36. Myles S, Chia JM, Hurwitz B, Simon C, Zhong GY, Buckler E, Ware D. Rapid genomic characterization of the genus *Vitis*. *PLoS One*. 2010;5:e8219.
37. Marrano A, Birolo G, Prazzoli ML, Lorenzi S, Valle G, Grandi MS. SNP-discovery by RAD-sequencing in a germplasm collection of wild and cultivated grapevines (*V. vinifera* L.). *PLoS One*. 2017;12:e1070655.
38. Mercati F, De Lorenzis G, Brancadoro L, Lupini A, Abenavoli MR, Barbagallo MG, Di Lorenzo R, Scienza A, Sunseri F. High-throughput 8K SNP array to assess genetic variability of the main grapevine cultivars from Sicily. *Tree Genet Genomes*. 2016;12:59.
39. De Lorenzis G, Chipashvili R, Failla O, Maghradze D. Study of genetic variability in *Vitis vinifera* L. germplasm by high-throughput Vitis8KSNP array: the case of Georgian genetic resources. *BMC Plant Biol*. 2015;5:154.
40. Degu A, Morcia C, Tumino G, Hochberg U, Toubiana D, Mattivi F, Schneider A, Bosca P, Cattivelli L, Terzi V, Fait A. Metabolite profiling elucidates communalities and differences in the polyphenol biosynthetic pathways of red and white Muscat genotypes. *Plant Physiol Biochem*. 2015;86:24–33.
41. De Lorenzis G, Squadrito M, Rossoni M, Simone Di Lorenzo G, Brancadoro L, Scienza A. Study of intra-variety diversity in biotypes of Aglianico and Muscat of Alexandria (*Vitis vinifera* L.) cultivars. *Aust J Grape Wine R*. 2017;23:32–42.
42. Ruffa P, Raimondi S, Boccacci P, Abbà S, Schneider A. The key role of Moscato bianco and Malvasia aromatica di Parma in the parentage of traditional aromatic grape varieties. *Tree Genet Genomes*. 2016;12:50.
43. Bowers JE, Dangl GS, Vignani R, Meredith CP. Isolation and characterization of new polymorphic simple sequence repeat loci in grape (*Vitis vinifera* L.). *Genome*. 1996;39:628–33.
44. Bowers JE, Dangl GS, Meredith CP. Development and characterization of additional microsatellite DNA markers for grape. *Am J Enol Vitic*. 1999;50:243–6.
45. Sefc KM, Regner F, Turetschek E, Glossl J, Steinkellner H. Identification of microsatellite sequences in *Vitis riparia* and their applicability for genotyping of different *Vitis* species. *Genome*. 1999;42:367–73.
46. Thomas MR, Scott NS. Microsatellite repeats in grapevine reveal DNA polymorphism when analysed as sequence-tagged sites (STSs). *Theor Appl Genet*. 1993;86:985–90.
47. Nei M. Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci U S A*. 1973;70:3321–3.
48. Xu S, Gupta S, Jin L. PEAS V.0: a package for elementary analysis of SNP data. *Mol Ecol Resour*. 2010;10:1085–8.
49. R Core Team: R: A language and environment for statistical Computing R Foundation for Statistical Computing 2004, Vienna, Austria. 2004. URL <http://www.r-project.org/>.
50. Kamvar ZN, Brooks JC, Grünwald NJ. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front Genet*. 2015;6:208.
51. Nei M. Genetic distance between populations. *Am Nat*. 1972;106:283–92.
52. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28:2731–9.
53. Caye K, Deist TM, Martins H, Michel O, Francois O. TESS3: fast inference of spatial population structure and genome scans for selection. *Mol Ecol Res*. 2016;16:540–8.
54. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet*. 2010;11:94.
55. Jombart T. Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*. 2008;24:1403–5.
56. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetic*. 1978;89:583–90.
57. Wright S. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*. 1965;19:395–420.
58. Goudet J. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Mol Ecol Notes*. 2005;5:84–6.
59. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet*. 2012;8:e1002967.
60. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population based linkage analyses. *Am J Hum Genet*. 2007;81:559–75.
61. Gu Z. Circize implements and enhances circular visualization in R. *Bioinformatics*. 2014;30:2811–2.
62. Butts C. Network: a package for managing relational data in R. *J Stat Softw*. 2008;24:1–36.
63. Jaillon O, Aury J-M, Noel B, Polcristi A, Clepet C, Casagrande A, et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*. 2007;449:463–7.
64. Bergamini C, Gasparro M, Sculli OV, Del Lungo S, Roccotelli S, Alba V, Perniola R, Antonacci D, Caputo AR: Recovery and valorization of the ancient viticultural biodiversity of Enotria in the modern Calabria Region. 40th World Congress of Vine and Wine, 29 May– 02 June 2017, Sofia, Bulgaria, Book of Abstract p. 81.
65. Schneider A, Raimondi S, Moreira F, De Santis D, Zappia R, Torello Marinoni D, Librandi N, Grandi MS. Contributo all'identificazione dei principali vitigni calabresi. *Frutticoltura*. 2009;1(2):46–55.
66. Lacombe T, Boursiquot JM, Laucou V, Di Vecchi-Staraz M, Pérois JP, This P. Large-scale parentage analysis in an extended set of grapevine cultivars (*Vitis vinifera* L.). *Theor Appl Genet*. 2013;126:401–14.
67. Calò A, Costacurta A, Maraš V, Meneghetti S, Crespan C. Molecular correlation of zinfandel (Primitivo) with Austrian, Croatian, and Hungarian cultivars and Kratošija, an Additional Synonym. *Am J Enol Vitic*. 2008;59:205–9.
68. Sefc KM, Lopes MS, Lefort F, Botta R, Roubelakis-Angelakis KA, Ibáñez J, Pejic I, Wagner HW, Glössl J, Steinkellner H. Microsatellite variability in grapevine cultivars from different European regions and evaluation of assignment testing to assess the geographic origin of cultivars. *Theor Appl Genet*. 2000;100:498–505.
69. Collesano G. Evoluzione varietale della viticoltura da vino in Sicilia. *Vignevini*. 1998;25:27–32.
70. Grassi F, Labra M, Imazio S, Spada A, Sgorbati S, Scienza A, Sala F. Evidence of a secondary grapevine domestication Centre detected by SSR analysis. *Theor Appl Genet*. 2003;107:1315–20.
71. McGovern PE: Ancient Wine: The Search for the Origins of Viniculture. Princeton, Princeton Univ press 2003.
72. Imazio S, Maghradze D, De Lorenzis G, Bacilieri R, Laucou V, This P, Scienza A, Failla O. From the cradle of grapevine domestication: molecular overview and description of Georgian grapevine (*Vitis vinifera* L.) germplasm. *Tree Genet Genomes*. 2013;9:641–58.
73. Riaz S, De Lorenzis G, Velasco D, Koehmstedt A, Maghradze D, Bobokashvili Z, Musayev M, Zdunic G, Laucou V, Andrew Walker M, Failla O, Preece JE, Aradhya M, Arroyo-García R. Genetic diversity analysis of cultivated and wild grapevine (*Vitis vinifera* L.) accessions around the Mediterranean basin and Central Asia. *BMC Plant Biol*. 2018;18:137.
74. Valamoti SM, Mangafa M, Koukoulis-Chrysanthaki C, Malamidou D. Grape-pressings from northern Greece: the earliest wine in the Aegean? *Antiquity*. 2007;81:54–61.
75. Valamoti SM. Harvesting the 'wild'? Exploring the context of fruit and nut exploitation at Neolithic Dikili Tash, with special reference to wine. *Veget Hist Archaeobot*. 2015;24:35–46.

76. Di Vora A, Castelletti L: 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* 1995, 176:333–358.
77. Olmo HP. The origin and domestication of the *Vinifera* grape. In: McGovern PE, editor. *The origins and ancient history of wine*. Amsterdam: Gordon and Breach; 1995. p. 31–43.
78. Arroyo-García R, Ruiz-García L, Bolling L, Ocete R, López MA, Arnold C, Ergul A, Söylemezoğlu G, Uzun HI, Cabello F, Ibáñez J, Aradhya MK, Atanassov A, Atanassov I, Balint S, Cenis JL, Costantini L, Goris-Lavets S, Grando MS, Klein BY, McGovern PE, Merdinoglu D, Pejic I, Pelsy F, Primikiriou N, Risovannaya V, Roubelakis-Angelakis KA, Snoussi H, Sotiri P, Tamhankar S, This P, Troshin L, Malpica JM, Lefort F, Martinez-Zapater JM. Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *sativa*) based on chloroplasts DNA polymorphism. *Mol Ecol*. 2006;15:3707–14.
79. De Andrés MT, Benito A, Pérez-Rivera G, Ocete R, Lopez MA, Gaforio L, Muñoz G, Cabello F, Martínez Zapater JM, Arroyo-García R. Genetic diversity of wild grapevine populations in Spain and their genetic relationships with cultivated grapevines. *Mol Ecol*. 2012;21:800–16.
80. Grassi F, De Mattia F, Zecca G, Sala F, Labra M. Historical isolation and quaternary range expansion of divergent lineages in wild grapevine. *Biol J Linn Soc*. 2008;95:611–9.
81. Mariani L, Parisi SG, Cola G, Failla O. Climate change in Europe and effects on thermal resources for crops. *Int J Biometeorol*. 2012;56:1123–34.
82. Francois O, Blum MG, Jakobsson M, Rosenberg NA. Demographic history of European populations of *Arabidopsis thaliana*. *PLoS Genet*. 2008;4:e1000075.
83. Hewitt GM. The genetic legacy of the quaternary ice ages. *Nature*. 2000;405:907–13.
84. Scienza A, Failla O. La circolazione dei vitigni in ambito Padano-Veneto ed Atesino: le fonti storico-letterarie e l'approccio biologico-molecolare. Trento: Istituto Trentino del Vino; 1996.
85. Bowers JE, Boursiquot JM, This P, Chu K, Johansson H, Meredith C. Historical genetics: the parentage of chardonnay, Gamay, and other wine grapes of northeastern France. *Science*. 1999;285:1562–5.
86. Vouillamoz JF, Monaco A, Costantini L, Stefanini M, Scienza A, Grando MS. The parentage of 'Sangiovese', the most important Italian wine grape. *Vitis*. 2007;46:19–22.
87. Marafioti G. *Croniche et antichità di Calabria*. Arnaldo Forni Editore: Anastatica; 1960.
88. Bergamini C, Perniola R, Cardone MF, Gasparro M, Pepe R, Caputo AR, Antonacci D. The molecular characterization by SSRs reveals a new south Italian kinship and the origin of the cultivar Uva di Troia. *Springerplus*. 2016;5:1562.
89. Scienza A. Per una storia della viticoltura campana. In: Scienza A, Boselli M, Napoli, Primi, editors. *Vini e vitigni della Campania, tremila anni di storia*; 2003. p. 5–60.
90. De Lorenzis G, Imazio S, Biagini B, Failla O, Scienza A. Pedigree reconstruction of the Italian grapevine Aglianico (*Vitis vinifera* L.) from Campania. *Mol Biotechnol*. 2013;54:634–42.
91. Crespan M, Cabello F, Giannetto S, Ibanez J, Kontic JK, Maletic E, Pejic I, Rodriguez-Torres I, Antonacci D. Malvasia delle Lipari, Malvasia di Sardegna, Greco di Gerace, Malvasia de Sitges and Malvasia dubrovacka—synonyms of an old and famous grape cultivar. *Vitis*. 2006;45:69–73.
92. Lacombe T, Boursiquot JM, Laucou V, Dechesne F, Vares D, This P. Relationships and genetic diversity within the accessions related to Malvasia held in the Domaine de vassal grape germplasm repository. *Am J Enol Viticult*. 2007;58:124–31.
93. Meneghetti S, Poljuha D, Frare E, Costacurta A, Morreale G, Bavaresco L, Calò A. Inter- and intra-varietal genetic variability in Malvasia cultivars. *Mol Biotechnol*. 2012;50:189–99.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

