



# Integrated Bayesian Approaches Shed Light on the Dissemination Routes of the Eurasian Grapevine Germplasm

## OPEN ACCESS

### Edited by:

Leila Riahi,  
University of Manouba, Tunisia

### Reviewed by:

Chograni Hnia,  
Ecole Supérieure d'Agriculture du  
Kef, Tunisia

Mériam Ben Romdhane,  
Institut Préparatoire aux Études  
Scientifiques et Techniques, Tunisia

### \*Correspondence:

Francesco Mercati  
francesco.mercati@ibbr.cnr.it  
Manna Crespan  
manna.crespan@crea.gov.it  
Francesco Sunseri  
francesco.sunseri@unirc.it

†These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Plant Systematics and Evolution,  
a section of the journal  
Frontiers in Plant Science

**Received:** 08 April 2021

**Accepted:** 07 July 2021

**Published:** 05 August 2021

### Citation:

Mercati F, De Lorenzis G, Mauceci A,  
Zerbo M, Brancadoro L, D'Onofrio C,  
Morcia C, Barbagallo MG, Bignami C,  
Gardiman M, de Palma L, Ruffa P,  
Novello V, Crespan M and Sunseri F  
(2021) Integrated Bayesian  
Approaches Shed Light on the  
Dissemination Routes of the Eurasian  
Grapevine Germplasm.  
Front. Plant Sci. 12:692661.  
doi: 10.3389/fpls.2021.692661

**Francesco Mercati<sup>1†</sup>, Gabriella De Lorenzis<sup>2†</sup>, Antonio Mauceci<sup>3</sup>, Marcello Zerbo<sup>1</sup>,  
Lucio Brancadoro<sup>2</sup>, Claudio D'Onofrio<sup>4</sup>, Caterina Morcia<sup>5</sup>, Maria Gabriella Barbagallo<sup>6</sup>,  
Cristina Bignami<sup>7</sup>, Massimo Gardiman<sup>8</sup>, Laura de Palma<sup>9</sup>, Paola Ruffa<sup>10,11</sup>,  
Vittorino Novello<sup>11</sup>, Manna Crespan<sup>8\*</sup> and Francesco Sunseri<sup>3\*</sup>**

<sup>1</sup> Istituto Bioscienze e Biorisorse, Consiglio Nazionale delle Ricerche, Palermo, Italy, <sup>2</sup> Dipartimento di Scienze Agrarie ed Ambientali, Università degli Studi di Milan, Milan, Italy, <sup>3</sup> Dipartimento Agraria, Università Mediterranea degli Studi di Reggio Calabria, Reggio Calabria, Italy, <sup>4</sup> Dipartimento di Scienze Agrarie, Alimentari e Agro-ambientali, Università degli Studi di Pisa, Pisa, Italy, <sup>5</sup> CREA - Centro di Ricerca per la Genomica e la Bioinformatica, Fiorenzuola d'Arda, Italy, <sup>6</sup> Dipartimento di Scienze Agrarie e Forestali, Università degli Studi di Palermo, Palermo, Italy, <sup>7</sup> Dipartimento di Scienze della Vita, Centro Biogest-Siteia, Università degli Studi di Modena e Reggio Emilia, Reggio Emilia, Italy, <sup>8</sup> CREA - Centro di Ricerca per la Viticoltura ed Enologia, Conegliano, Italy, <sup>9</sup> Dipartimento di Scienze Agrarie, Alimenti, Risorse Naturali e Ingegneria, Università degli Studi di Foggia, Foggia, Italy, <sup>10</sup> Istituto per la Protezione Sostenibile delle Piante, Consiglio Nazionale delle Ricerche, Torino, Italy, <sup>11</sup> Dipartimento di Scienze Agrarie, Forestali e Alimentari, Università degli Studi di Torino, Grugliasco, Italy

The domestication and spreading of grapevine as well as the gene flow history had been described in many studies. We used a high-quality 7k SNP dataset of 1,038 Eurasian grape varieties with unique profiles to assess the population genetic diversity, structure, and relatedness, and to infer the most likely migration events. Comparisons of putative scenarios of gene flow throughout Europe from Caucasus helped to fit the more reliable migration routes around the Mediterranean Basin. Approximate Bayesian computation (ABC) approach made possible to provide a response to several questions so far remaining unsolved. Firstly, the assessment of genetic diversity and population structure within a well-covered dataset of ancient Italian varieties suggested the different histories between the Northern and Southern Italian grapevines. Moreover, Italian genotypes were shown to be distinguishable from all the other Eurasian populations for the first time. The entire Eurasian panel confirmed the east-to-west gene flow, highlighting the Greek role as a “bridge” between the Western and Eastern Eurasia. Portuguese germplasm showed a greater proximity to French varieties than the Spanish ones, thus being the main route for gene flow from Iberian Peninsula to Central Europe. Our findings reconciled genetic and archaeological data for one of the most cultivated and fascinating crops in the world.

**Keywords:** *Vitis vinifera* L. subsp. *sativa*, SNP array, genetic structure, LD decay, ancestry coefficients, migration events

## INTRODUCTION

Grapevine (*Vitis vinifera* L. subsp. *sativa*) is one of the earliest domesticated and most cultivated crops worldwide, having a great impact on the agri-food economy and prized for its fruits and wines. Europe is the largest producer (primarily Italy, Spain, and France), followed by Asia and America. Nowadays, worldwide grape production reaches nearly 78 M tons (<http://faostat.fao.org/>).

Historical and archaeological evidence dated grapevine domestication back to the Neolithic age (ca. 8,500–4,000 BC), when human populations began to collect and propagate *Vitis* forms to improve fruit and wine production (McGovern et al., 1986; Zohary and Hopf, 2000). Domestication took place in the region between the Caucasus and Mesopotamia (Georgia, Iran, Turkey), domesticating wild populations of *Vitis vinifera* subsp. *silvestris*, considered to be the ancestor of the cultivated form, the subsp. *sativa* (Levadoux, 1956; Myles et al., 2011; De Andrés et al., 2012; McGovern et al., 2017). From the oldest domestication sites, cultivated grapevines disseminated westward into neighboring regions (Egypt and Lower Mesopotamia), reaching the Mediterranean Basin, together with the development of human culture. Romans later spread grapevine cultivation in the temperate areas of Europe, following the main trade routes along the major rivers Rhine, Rhone, and Danube. Islam then played an important role in spreading table grapes to Northern Africa, Spain, and Middle East (This et al., 2006).

The Italian Peninsula seems to have had a key role in the spreading of grapevine and viticulture practices from Greece to Central and Western Europe, due to its strategic position on the Mediterranean Sea. In the eighth century BC, the establishment of Greek colonies in the Southern Italy was the main driver of viticulture expansion and development in these areas. In the beginning, the settlers conceivably introduced vines from their places of origin, nonetheless they plausibly also cultivated local populations crossed with Greek varieties (Buono and Vallariello, 2002; McGovern, 2003; Scienza, 2004; Marvelli et al., 2013).

In Europe, a long period of significant successes for viticulture has been witnessed due to the human selection of new and valuable cultivars, up to the current age. To date, the *Vitis* International Variety Catalog (<http://www.vivc.de>) and the Italian Register (<http://catalogoviti.politicheagricole.it>) list more than 13,000 and nearly 800 Eurasian grapevine varieties, respectively.

Genetic diversity and population structure of the cultivated varieties has been extensively assessed throughout highly polymorphic microsatellite markers (simple sequence repeat, SSR) (Cipriani et al., 2010; Laucou et al., 2011; Lacombe et al., 2012; Bacilieri et al., 2013; Emanuelli et al., 2013). Recently, by advanced next-generation sequencing, three high-throughput *Vitis* genotyping chip arrays, including 9, 18, and 37k single nucleotide polymorphism (SNP) loci were developed (Myles et al., 2010; Marrano et al., 2017; Laucou et al., 2018). These new flexible and performing tools

became as valuable as SSR in providing grapevine genetic diversity, structure information, pedigree analysis, gene flow, and genome-wide association analysis (GWAS) with higher-throughput and cost-effectiveness than SSRs (Myles et al., 2011; De Lorenzis et al., 2015, 2019; Mercati et al., 2016; Marrano et al., 2017; Laucou et al., 2018; Sunseri et al., 2018).

Despite the large efforts to analyze the genetic variability is available worldwide, a myriad of questions remains unsolved. First, the relationships within and among populations: What is the spreading history of Italian and the other Eurasian populations? What was the role of Caucasian germplasm in shaping the grapevine populations cultivated worldwide? How much did the ancient population affect the shape of the present ones? Does the grapevine genetic variability fit the human migration routes around the Mediterranean Basin?

To address the above questions, a comprehensive study on 384 varieties belonging to the Italian grapevine germplasm and 654 additional varieties belonging to Eurasia were used to unravel the complexity of grapevine germplasm relationships.

The entire panel allowed us to improve the knowledge on genetic diversity among grapevine populations and their dissemination routes by assessing the genetic structure and population relatedness, computing linkage disequilibrium (LD) decay, and ancestry coefficients, as well as, inferring migration events through the evaluation of putative scenarios.

## MATERIALS AND METHODS

### Sample Collection and Genotype Calling

Three-hundred-eighty and four grapevines (CREA—<https://www.crea.gov.it>—and AGER consortium—<https://www.progettoager.it>) showing unique SNP profiles and covering most of the Italian germplasm were selected from 615 genotypes firstly investigated for parentage analysis (D'Onofrio et al., 2021). To provide an overview of grapevine dissemination throughout Eurasia, 654 varieties from the available datasets (De Lorenzis et al., 2015, 2019; Laucou et al., 2018) were included (**Supplementary Table 1**). All the genotypes were characterized using the 18k SNP array (Laucou et al., 2018). A stringent filtering [SNP NA rate > 1% and minor allele frequency (MAF) < 5%; NA rate for each individual > 5%; see D'Onofrio et al., 2021] was applied to ultimately obtain an SNP panel (**Supplementary Table 1**), then the duplicated profiles detected by calculating the pairwise percentage of mismatches between (20% was used as cut-off) individuals were deleted. After filtering, 6,770 high-quality SNPs and 1,038 unique genotypes were retained (**Table 1** and **Supplementary Table 1**). Based on the previous information (Bacilieri et al., 2013; Laucou et al., 2018; De Lorenzis et al., 2019), 1,038 genotypes were distinguished in six groups for geographic origin: (i) Balkans (BALK); (ii) Eastern Mediterranean and Caucasus, Middle and Far East, Russia and Ukraine (EMCA-MFEAS-RUUK); (iii) Iberian Peninsula (IBER); (iv) ITAP-Northern and Center Italy (north-center); (v) ITAP-Southern Italy (south); (vi) Western and Central Europe (WCEUR).

**TABLE 1** | Grapevine varieties collected across Europe, Caucasus, Middle, and Far East, genotyped by 18k SNP genotyping array and arranged based on their passport data (for a detailed list, see **Supplementary Table 1**).

Group code	Origin	Subgroup name	Countries	Samples' number
BALK	Balkans	Eastern Europe	ALB, BGR, HUN, RO	125
		Balkan Peninsula	BIH, CYP, GRC, HRV, MNE, SCG, SVN	
EMCA-MFEAS-RUUK	Eastern Mediterranean and Caucasus	Caucasus and Turkey	AFG, ARM, AZE, GEO, TJK, TKM, TUR, UZB	167
		Near East	ISR, LBN, SYR	
	Middle and Far East	Middle East	IRN	
	Russia and Ukraine	Russia and Ukraine	MDA, RUS, UKR	
IBER	Iberian Peninsula	Iberian Peninsula	ESP, PRT	196
ITAP-north-center	Italian Peninsula	Italian Peninsula—North	ITA	143
		Italian Peninsula—Center	ITA	105
ITAP-south		Italian Peninsula—South	ITA	136
WCEUR	Western and Central Europe	Western Europe	FRA	166
		Central Europe	AUT, DEU, CHE, TCH	
<i>Total</i>				1,038

### Genetic Diversity and Population Structure Cluster and Principal Coordinates Analyses

Phylogenetic analysis and principal coordinates analysis (PCoA) on both the Italian (384 genotypes) and whole (1,038) grapevine dataset were performed to estimate the overall relationship among the varieties. The distance-based dendrogram with Nei's genetic distance (Nei, 1972), UPGMA algorithm (bootstraps based on 1,000 re-samplings), and PCoA were developed using R/poppr (Kamvar et al., 2015) and R/adeigenet packages (Jombart, 2008), respectively.

### Structure and Discriminant Analysis of Principal Components

The number of genetic pools (*K*) was computed for both datasets using an admixture model performed through fastSTRUCTURE (Raj et al., 2014), using the input files (.bed, .bim, and .fam) generated by PLINKv1.07 (Purcell et al., 2007) and the *structure.py* python script. The best number of model complexity (*K*) was chosen applying the algorithm for multiple choices, *chooseK.py*. To extract the optimum *K*, the prediction error for each *K* was computed by the cross validation (CV) function. Bayesian analyses for both datasets were performed also by discriminant analysis of principal components (DAPC) implemented in the R/adeigenet. The Bayesian information criterion (BIC) method (Neath and Cavanaugh, 2012) was used to infer the *K-means* clustering. The analyses were performed independent of geographic origin.

### Wright's Fixation Index and Neighbor-Joining Tree

The Wright's fixation index (*Fst*), a population pairwise index (Wright, 1965), was computed to evaluate the genetic distances among the six groups outlined in **Table 1** using R/HierFstat (Goudet, 2005). A neighbor-joining (NJ)-tree based on *Fst*-values was developed with R/adeigenet.

### Identity by Descent Analysis and Linkage Disequilibrium Decay

To identify the common ancestry of putative recent, identity by descent (IBD) segments were detected in the whole dataset. Relatedness analysis using IBD estimation was conducted through R/SNPrelate (Zheng et al., 2012) using the Method of Moments (MoM) (Purcell et al., 2007), and a multi-dimensional scaling (MDS) analysis was performed on the *n* × *n* matrix of genome-wide IBD pairwise distances. The shared haplotypes between individuals with unknown relationships were also evaluated, using PLINK with the default setting. The average pairwise IBD between chromosomes from different individuals in the groups was summarized by comparison in a heatmap.

The estimation of LD as the Pearson's squared correlation coefficient (*r*<sup>2</sup>) was calculated between each pair of molecular markers (Zhao et al., 2005) for the six groups (**Table 1**). To avoid distortions, a representative and equivalent number of samples for each group was defined by the R/corehunter (Thachuk et al., 2009), using BALK population size (125 varieties) as the cutoff. The pairwise LD, *r*<sup>2</sup> was calculated using PLINK and the parameters, such as *-maf 0.05*, *ld-window-r2 set to zero*, *ld-window 99999*, and *ld-window-kb 10000* (Hill and Robertson, 1968). Bins of 100 kb among all pairwise combinations were determined based on the physical distance of each SNP pair.

### Inference of Migration Events

The maximum likelihood (ML)-tree of data collected and a gene flow model among the geographic groups (**Table 1**) were developed by TreeMix (Pickrell and Pritchard, 2012). Stratified allele frequencies from PLINK were converted into the TreeMix format using the *plink2treemix.py* script and used as input (<https://speciationgenomics.github.io/Treemix/>; Pickrell and Pritchard, 2012). Forty independent ML searches following the procedure described by Zecca et al. (2020) were performed. The results were filtered based on their likelihood values using the R/cfTrees (Zecca et al., 2020), duplicates were deleted, and the

best-scoring ML tree was used. The gene flow model among the groups was investigated through migration events ( $m$ ). Migration edges were tested 10 times from 1 to 5 with different random seeds each time and using blocks ( $k$ ) of 20 SNPs, to check for convergence in terms of the likelihood value of each model, and the variance explained in each migration event was added. Standard errors (SE) and bootstrap replicates (bootstrap) were used to assess the confidence in the inferred tree topology and the weight of migration events. To automate the choice of the best migration event, an *ad hoc* statistic based on the second-order rate of change in the likelihood weighted by SD was adopted through R/OptM (Fitak, 2019). However, since the true model was considered when the migration edges ( $m$ ) explained 99.8% of the variance in ancestry between groups, only the model showing this cutoff was believed to be the best one (Pickrell and Pritchard, 2012). In addition, only the runs with all statistically significant incorporated migration edges were considered. The residuals from the fitted models chosen for our data were visualized using the R script *plot\_resid*.

## Inference of Spatial Population Structure Through Ancestry Coefficient

To assess the population structure and the putative different grapevine spreading routes, a spatial ancestry estimation was performed using R/TESS3 (Caye et al., 2016). The ancestry coefficient was evaluated on the whole dataset by adding 21 genotypes from the Maghreb, named MAGH (Laucou et al., 2018), reaching 1,059 varieties (Supplementary Table 1). The samples from Northern Africa appeared useful to verify their possible role as a bridge between the first domestication center, EMCA-MFEAS-RUUK, and IBER. The MAGH group was not included in the previous analysis due to its small size compared to the other six groups (Supplementary Table 1). The number of the best  $K$  was chosen after a cross-entropy criterion evaluation for each  $K$  (Frichot et al., 2014; Caye et al., 2016). The values of Q-matrix for the best  $K$  were interpolated on a geographic map.

## Approximate Bayesian Computation Analysis

An approximate Bayesian computation (ABC) approach implemented with DIY-ABC 2.0 software was adopted to establish the most likely grapevine gene flow scenarios (Cornuet et al., 2014). The genotypes belonging to the six groups (Table 1) and Maghreb were used (Supplementary Table 1). Four groups of hypotheses were evaluated.

### About the Origin of Southern Italian Genotypes

The first hypothesis aimed to clarify the origin of Southern Italy germplasm, assuming that grapevine was spread from Greek shores to the Southern Italy and then to the Central and Northern Italy (Buono and Vallariello, 2002; Scienza, 2004; Myles et al., 2011; Bacilieri et al., 2013; Riaz et al., 2018; De Lorenzis et al., 2019). The three tested scenarios are as follows: (i) Scenario 1 assumed a flow of domesticated genotypes from BALK into ITAP-south, and then into ITAP-north-center; (ii) Scenario 2 assumed that domesticated grapevines from BALK first flowed into ITAP-north-center, then to ITAP-south; (iii) Scenario 3

assumed that ITAP-south and ITAP-north-center germplasm derived independently from BALK.

### The Gene Flow From Italy to the Western and Central Europe

The second hypothesis took into account the gene flow from Italy to the Western and Central Europe to dissect the relationship among the Northern Italian and the Western-Central European germplasm (De Lorenzis et al., 2019). Three scenarios were depicted: (i) Scenario 4 assumed that ITAP-south germplasm spread in the Northern-Central Italy and then WCEUR; (ii) Scenario 5 assumed that ITAP-south germplasm spread first in WCEUR and then in ITAP-north-center; (iii) Scenario 6 assumed that an admixture event between ITAP-south and WCEUR genotypes flowed into ITAP-north-center germplasm.

### The Gene Flow From ITAP-South to WCEUR Through IBER

The third case proposed the hypothesis of a gene flow from the Southern Italy to WCEUR *via* IBER (Buono and Vallariello, 2002); The three supposed scenarios are as follows: (i) Scenario 7 assumed that grapevine germplasm from ITAP-south was introduced first into IBER and then into WCEUR; (ii) Scenario 8 assumed that WCEUR genotypes derived from an admixture event between ITAP-south and IBER genotypes; (iii) Scenario 9 assumed the contrary hypothesis of Scenario 7.

### The Gene Flow From Middle and Far East to IBER Through Northern Africa

Lastly, the hypothesis of a gene flow from the Middle and Far East to IBER through Northern Africa was verified. The three scenarios were as follows: (i) Scenario 10 assumed a gene flow from EMCA-MFEAS-RUUK to Northern Africa (MAGH) and then to IBER (ii) Scenario 11 assumed that an admixture event between MAGH and EMCA-MFEAS-RUUK germplasm flowed into the IBER genotypes; (iii) Scenario 12 assumed a contrary hypothesis to Scenario 10.

### Calculation Setting Up, Core Collections, and Statistics

To reduce the computational power required to perform the ABC analysis, the dataset were pruned with PLINK, removing SNP loci pairs not in optimal LD, using a sliding window of 50 bp and a step size of 5 SNPs, with a variance inflation factor 2 (Orozco-terWengel et al., 2015). A core collection was also built for each population using the allele coverage allocation strategy implemented in R/corehunter, which are useful to maximize the allele proportion. One hundred thousand datasets were simulated by do it yourself (DIY)-ABC, for a total of  $12 \times 10^5$  simulated datasets, considering all scenarios. Mean gene diversity across all loci, mean across loci of *Fst* distances between pairs of geographic groups, and mean across loci of *Nei's* genetic distances between pairs of geographic groups were calculated. The most appropriate scenario for each hypothesis was selected comparing the summary statistics of simulated and observed datasets. The posterior probabilities [with 95% confidence interval (CI)] of each competing scenario were also estimated following a logistic



regression approach on the 1% of simulated datasets closest to the observed dataset. The ability of ABC approach to discriminate between scenarios (ABC performance) was evaluated. Per each scenario, Type I and II error probabilities were also estimated (Cornuet et al., 2014).

Lastly, to validate the goodness-of-fit of the selected scenario, a model checking test was performed, carrying out the local linear regression on the 1% of closest simulated datasets, after adopting a logit transformation to parameters, as suggested by the DIY-ABC user manual (Cornuet et al., 2014).

## RESULTS

### Italian Germplasm Genetic Diversity Phylogenetic and PCoA Analyses

Phylogenetic analysis among the Italian genotypes highlighted two main groups showing the distinction between Southern and Northern Italian (ITAP-south and -north) varieties, except for 14% of ITAP-south that clustered in the Northern Italian group and *vice versa* for about 16% of the samples (Figure 1A). The Central Italian (ITAP-center) samples were divided between Northern (57%) and Southern (43%) clusters (Figure 1A). A PCoA analysis confirmed the genetic relationships within the Italian germplasm (Supplementary Figure 1), splitting ITAP-north and ITAP-south genotypes into two groups, with the ITAP-center samples divided between them.

### Structure and DAP Analysis

The optimum  $K$  was firstly evaluated with fastSTRUCTURE, providing an optimal value between 2 and 6 (Model complexity that maximizes the marginal likelihood = 6; Model components used to explain structure in data = 2). Based on the prediction error, the lowest model complexity within the range was explained at  $K = 5$  (Supplementary Figure 2A).

The DAPC analysis was then performed to further investigate the Italian germplasm and to confirm the group assignments.  $K$ -means clustering was estimated using the BIC score, defining the best cluster number at  $K = 3$  (Supplementary Figure 3 and Supplementary Table 2). The scatter plot highlighted an ITAP-north varieties cluster (mainly gathered in Pool 1, red) and two clusters (Pools 2 and 3, blue and green) including ITAP-center and ITAP-south genotypes, respectively (Figure 1B). The posterior probability at  $K = 3$  underlined that nearly all ITAP-south genotypes shared both Pools 2 and 3; the ITAP-center samples split in Pools 1 and 3, while ITAP-north group mainly forms Pool 1 (Figure 1C).

In detail, 196 genotypes (51%) were attributable to a specific pool with a high percentage (ancestry membership >70%) by fastSTRUCTURE. About sixty-two (87%) out of 71 ITAP-north samples formed Pool 1, the ITAP-center genotypes were mainly split between Pools 1 (28) and 3 (19), while Pool 2 is formed almost only by the ITAP-south genotypes (88%), and four from ITAP-center (Supplementary Table 3). About 55% of ITAP-south genotypes were characterized by Pool 3, appearing as a bridge among the Italian varieties (Figure 1B and Supplementary Table 3).

In agreement, DAPC analysis assigned 378 out of 384 (98.4%) Italian varieties to one of the three major pools, setting a cut-off of 70% posterior membership probabilities (Supplementary Table 3). Almost all the Italian varieties were assigned to a pool, with the highest percentage (99%) for both the ITAP-north (141 out of 143) and ITAP-south (135 out of 136) groups, according to Bayesian analysis (Figure 1C and Supplementary Table 3). Although the best model was fixed at  $K = 3$  (Figure 1), the same trend in the distribution of Northern, Central, and Southern Italian genotypes was also shown at  $K = 5$ , based on the cross-validation function (Supplementary Figure 2).

The DAPC results matched with that of PCoA and the unweighted pair group method with arithmetic mean (UPGMA). According to these results, the Northern and Central Italian genotypes were joined and named ITAP-north-center (Supplementary Table 1).

### Genetic Relationship Among Eurasian Grapevine Populations by the Whole Dataset

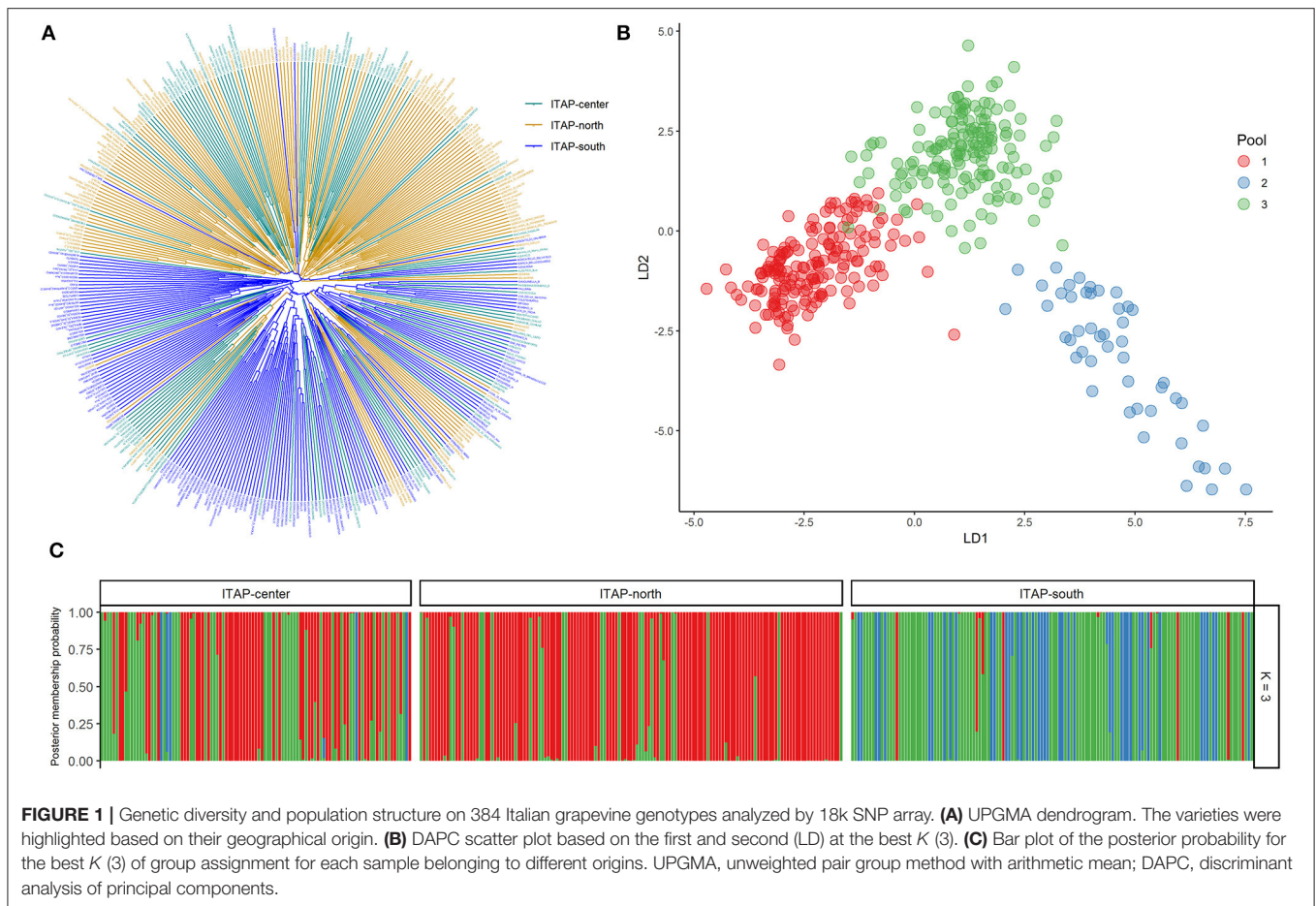
#### Unweighted Pair Group Method With Arithmetic Mean and PCoA Clustering

Cluster analysis distinguished genotypes from the IBER, ITAP-north-center, and WCEUR in three clusters (Figure 2A). Nearly 50 Iberian samples (~25%) clustered with WCEUR genotypes. The ITAP-south, BALK, and EMCA-MFEAS-RUUK samples were grouped in a major assorted cluster, where a sub-cluster of ITAP-north-center genotypes was included (Figure 2A, lower part). The ITAP-south varieties appeared more closely related to BALK than EMCA-MFEAS-RUUK genotypes. Another significant cluster included EMCA-MFEAS-RUUK and ITAP-south genotypes, in the upper part of the dendrogram, as a bridge between IBER and WCEUR populations (Figure 2A).

A similar distribution was obtained through PCoA analysis (Supplementary Figure 4), with the whole set divided into three main groups, with some expected overlapping: ITAP-north-center samples grouped together with WCEUR genotypes and separated by Principal Coordinate (PCo) 1; EMCA-MFEAS-RUUK samples clustered together with BALK and ITAP-south groups resulting in complete separation from IBER varieties by PCo2 (Supplementary Figure 4).

### Structure and DAPC

The optimum pool number was recorded at  $K = 9$ , using both prediction error and BIC method (Supplementary Table 2 and Supplementary Figure 5). Each population showed the prominent membership (>0.4) to an owner pool computed through both fastStructure and DAPC, except WCEUR (Supplementary Table 4). Moreover, Pool 5 is not owned by any population, with the highest membership related to Moscato bianco (>0.99), the ancestor of muscat flavored varieties (Cipriani et al., 2010; Ruffa et al., 2016). Pool 5 grouped varieties belonging to different populations, showing



first- or second-degree relatedness with Moscato bianco (**Supplementary Table 1**).

The DAPC scatter plot showed a triangle-shaped distribution, highlighting that Pool 9 is divergent from the others (**Figure 2B**). This pool mainly encompassed genotypes belonging to the EMCA-MFEAS-RUUK population (58%, 98 out of 167 genotypes), while the remaining 30% showed the highest membership with Pool 3. Pools 2 and 3 encompassed the BALK germplasm with 60% (75 out of 125) and 20%, respectively (**Figure 2** and **Supplementary Table 4**). The ITAP-south population showed a unique genetic structure (73%, 100 out of 136) in Pool 6, according to DAPC results (**Figure 2C** and **Supplementary Table 1**). All the other populations were split into at least two main pools: ITAP-north-center genotypes shared Pools 6 and 7, IBER population belonged to Pools 5 and 8, while WCEUR genotypes encompassed Pools 1, 4, and 5, the last in common with IBER (**Figure 2** and **Supplementary Table 4**).

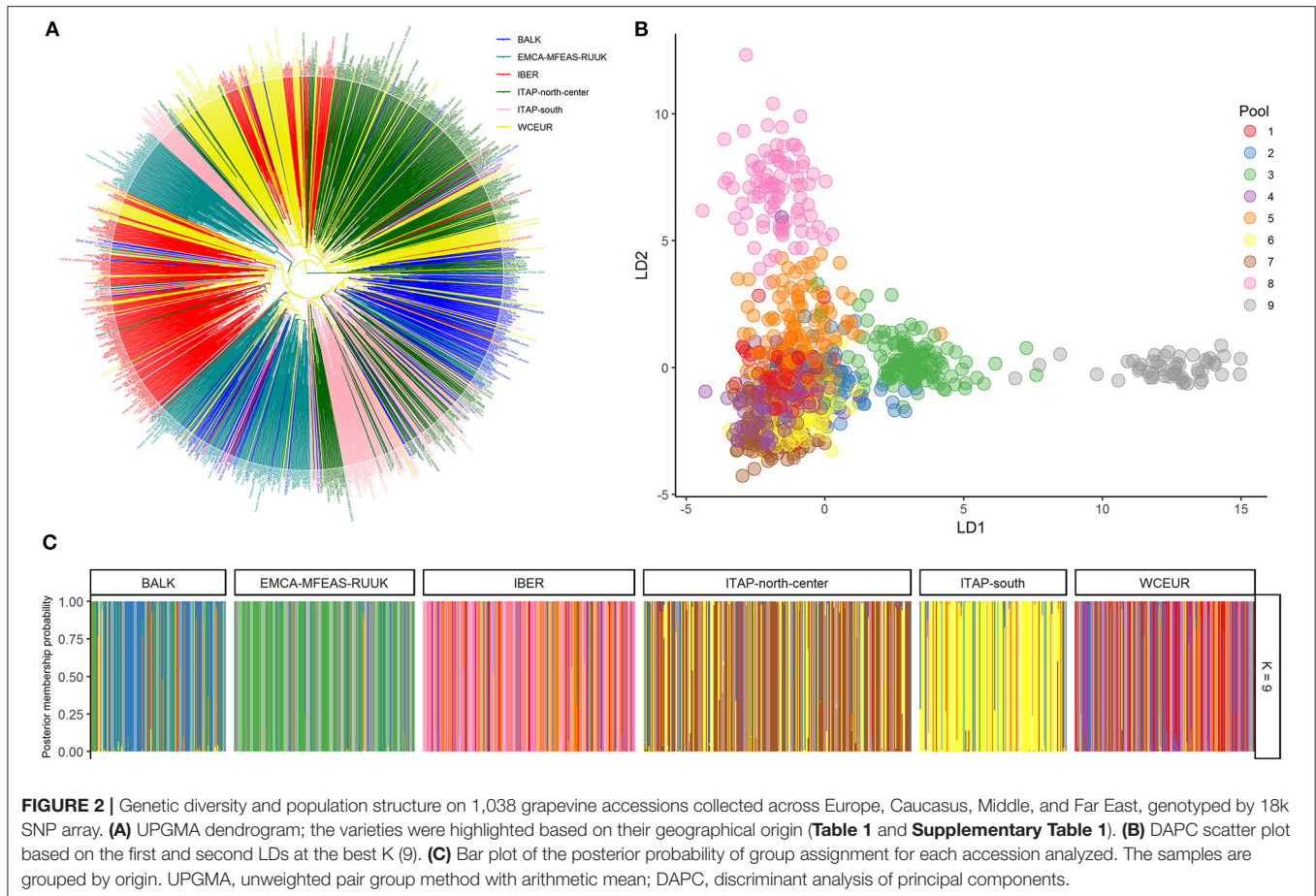
### Fixation Index, IBD, and LD Decay

The EMCA-MFEAS-RUUK and WCEUR germplasm showed the highest genetic distance according to pairwise  $F_{st}$ -values (0.033; **Supplementary Table 5**). Large genetic divergences were observed between EMCA-MFEAS-RUUK and IBER, as well as ITAP-north-center groups ( $F_{st} = 0.024$ ). The lowest  $F_{st}$

pairwise-value (0.009) was recorded between WCEUR and ITAP-north-center groups, followed by EMCA-MFEAS-RUUK and BALK groups (0.011) and ITAP-south and BALK (0.011) (**Supplementary Table 5**). The NJ tree developed by  $F_{st}$ -values (**Figure 3A**) split up all the populations showing a structure consistent with previous results. The EMCA-MFEAS-RUUK and BALK groups formed one branch and differed greatly from WCEUR and ITAP-north-center groups, both included in another main branch. The IBER and ITAP-south populations are separated, with the last closely related to BALK samples, in agreement with PCoA (**Supplementary Figure 4**).

Based on genome-wide IBD pairwise distances, the MDS analysis showed a distinctness of IBER (red), as well as EMCA-MFEAS-RUUK (cyan), and BALK (blue) populations, while the other populations were not clearly separated (**Figure 3B**). The average amount of IBD shared between individuals in the six populations underlined the relatedness between BALK and ITAP-south genotypes, as well as between ITAP-north-center and WCEUR genotypes, while IBER and EMCA-MFEAS-RUUK populations showed the most consistent genetic distances (**Figure 3C**).

The average LD declined with the increase of physical distance between markers and appeared fast in all the populations. The fastest decay was observed in EMCA-MFEAS-RUUK population,



followed by BALK and in the pair from ITAP (Figure 3D). In contrast, a relatively slower LD decay was recorded for WCEUR and IBER. Thus, the LD decay rate decreased with the increase of geographic distance from the Middle and Far East to the Iberian Peninsula, in an East-to-West gradient. The average distance at which the LD-value reached 0.05, varied among populations, indeed the distance accounted for  $\sim 3.2$  Mb in the varieties belonging to IBER population, while it decreased to  $\sim 1$  Mb and 1.8 Mb in the EMCA-MFEAS-RUUK and BALK populations, respectively (Figure 3D).

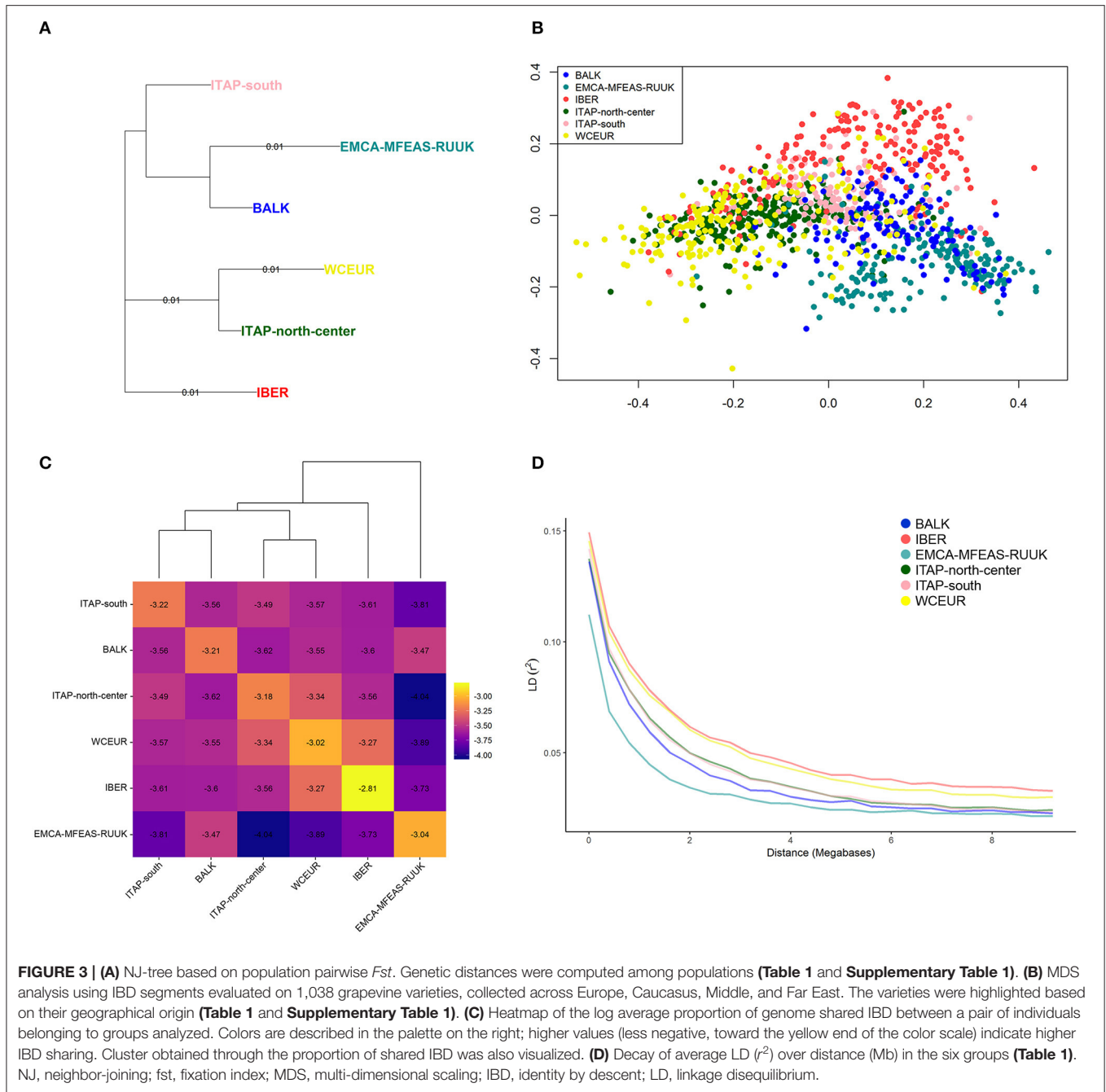
## Gene Flow Evaluation Through the Estimation of Possible Migration Events

To assess the complex demographic history of the cultivated grapevine, a migration map of ancestral populations based on gene flow was drawn (Figure 4). Forty preliminary ML searches showed the same likelihood ( $\ln$ )-value, with a similar profile in terms of topology and branch lengths; therefore one ML was chosen randomly. The ML-tree explained 98% of the variance in relatedness between grape populations from different geographic origins, supporting the tree-like history, with branch lengths proportional to the drift amount from the group split (Figure 4A; Supplementary Figure 6A; and

Supplementary Table 6). Furthermore, the residuals from the fit of the selected model highlighted a not complete explanation of different population ancestry (Figure 4B). Therefore, five migration events ( $m$ ) were added in sequence, showing the saturation of the model likelihood at four additional migration edges (99.8% of variance explained; Supplementary Figure 6A). Overall, each run at  $m = 4$  showed an improvement over the tree without migration, and the model with the highest significance level ( $p < 0.001$ ) for all migration edges recorded was chosen with the best fits at four inferred migration events (Supplementary Table 7). These events were from ITAP-south to ITAP-north-center (weight 44%,  $p < 0.001$ ), from EMCA-MFEAS-RUUK to IBER (weight 27%;  $p < 0.001$ ), from IBER to WCEUR (weight 18%;  $p < 0.001$ ), and from ITAP-north-center to EMCA-MFEAS-RUUK (weight 4%;  $p < 0.001$ ) groups (Figure 4C).

Finally, the optimal number of  $m$  was also verified through the Evanno and linear methods, inferring two admixture events from ITAP-south to ITAP-north-center (weight 44%;  $p < 0.001$ ) and from EMCA-MFEAS-RUUK to IBER (weight 35%;  $p < 0.001$ ) populations (Supplementary Figures 6B,C, 7A and Supplementary Table 6). Comparing both explained variances and residual fit between the two possible models, four migrations or two admixture events, the best one of  $m = 4$  with the





highest  $f$  (0.998 vs. 0.992) and the lowest residual values (SE  $\pm 2.7$  vs. SE  $\pm 6.8$ ) (Figure 4D; Supplementary Figure 7B; and Supplementary Table 6) were obtained.

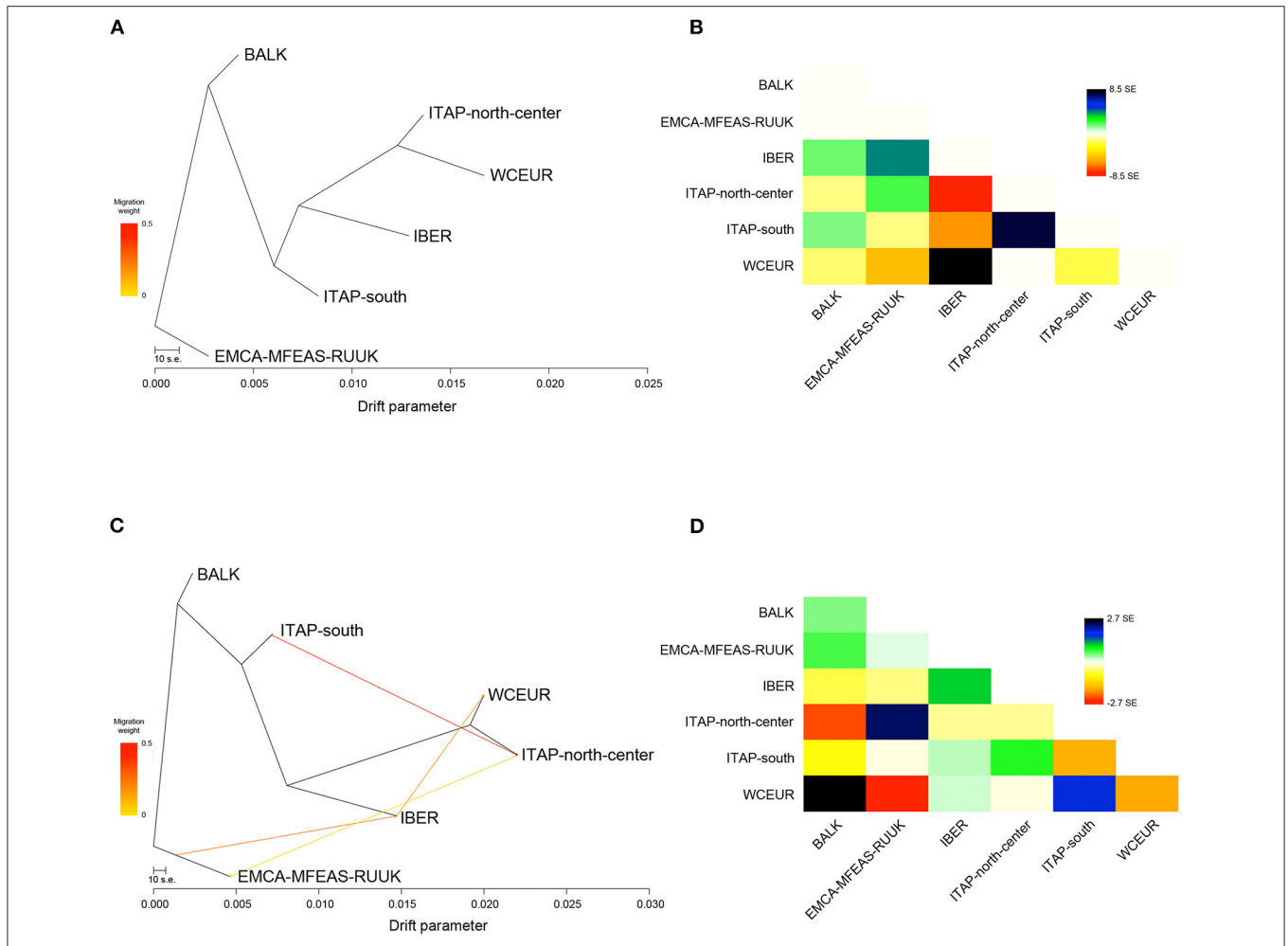
### Grapevine Spreading Evaluation Through Ancestry Coefficient Analysis

The ancestry estimation was performed adding 21 Northern African (MAGH) genotypes (Laucou et al., 2018) to the whole dataset to better overview the conceivable grapevine migration routes.

The analysis of ancestry through the cross-entropy criterion on the total 1,059 varieties estimated a single fixed  $K = 8$  (Supplementary Figure 8), confirming the distinctness among grapevine populations as revealed above (Figure 5A).

The ancestry coefficients interpolated to the geographic coordinates clearly showed the grapevine spreading from the first domestication center (Caucasus) to the Mediterranean Basin (Figure 5A). Moreover, four out of the seven grapevine populations resulted in the ancestors of a specific pool, except ITAP-north-center, WCEUR and MAGH populations, showing an admixed structure (Figure 5B and Supplementary Table 8).





**FIGURE 4 | (A)** The ML-tree inferred by TreeMix. The scale bar shows 10 times the average SE of the entries in the sample covariance matrix. Drift parameter is shown on the x-axis. **(B)** Residual fit from ML-tree with SE ± 8.5. High standard errors suggest strong candidate groups for admixture events. Positive residuals characterize groups where the model underestimates the observed covariance; similarly, pairs of groups where the model overestimates the observed covariance showed negative residuals. Colors are described in the palette on the right. **(C)** ML-tree that best fits the data with 4 additional migration edges: EMCA-MFEAS-RUUK to IBER ( $p < 0.001$ ), ITAP-south to ITAP-north-center ( $p < 0.001$ ), IBER to WCEUR ( $p < 0.001$ ), and ITAP-north-center to EMCA-MFEAS-RUUK ( $p < 0.001$ ). Migration arrow is colored according to its weight and colors are explained in the palette on the left. **(D)** Residual fit from ML-tree at  $m = 4$ , with SE ± 2.7. ML, maximum likelihood; SE, standard error.

### The Best Scenarios for East-To-West Grapevine Migration Routes

Four groups of scenarios were tested using ABC approach to define the possible East-to-West grapevine migration routes (Figure 6 and Supplementary Figure 9). The ABC approach was performed on a pruned SNP dataset (2k) and a core collection per each population, to maintain the available genetic diversity: 21, 31, 30, 33, 34, 37, and 32 genotypes were chosen for MAGH, BALK, EMCA-MFEAS-RUUK, IBER, ITAP-north-center, ITAP-south, and WCEUR groups, respectively.

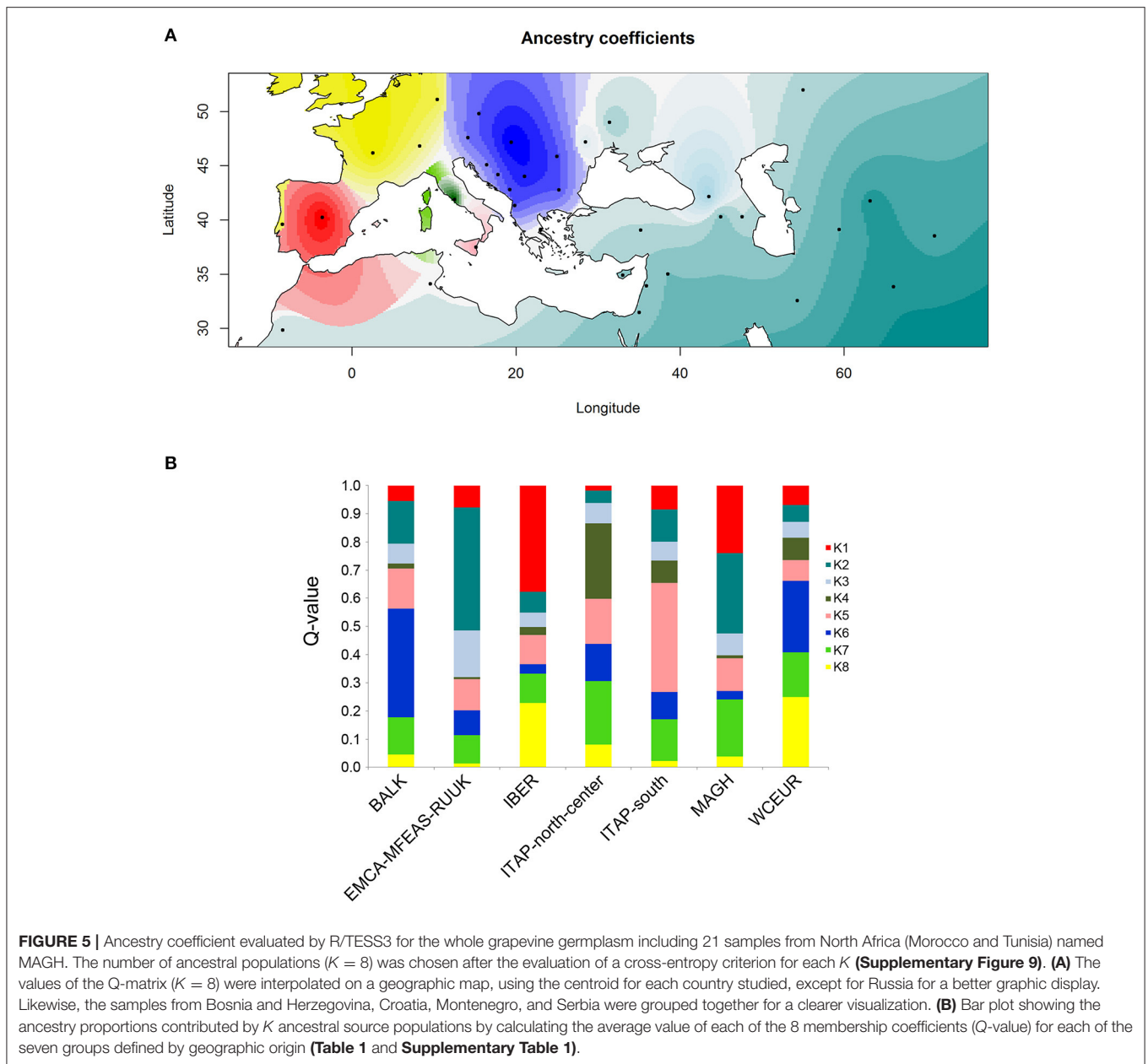
### The Origin of ITAP-South Germplasm

Scenario 1 showed the best fitting values, assuming a gene flow from BALK to ITAP-south, then to ITAP-north-center (Figure 6A). For Scenario 1: (i) the highest logistic regression values were recorded (Supplementary Figure 9A); (ii) the

posterior probability (Pp)-value was much higher than the others, showing a mean around 74%, with a 95% CI (72–79%) not overlapped with those of the other scenarios; (iii) the lowest number (3 out of 9) of statistics deviated significantly from the observed values (Table 2 and Supplementary Table 9). The ABC performance was of 77% and Type I and Type II error rates were <20% (Supplementary Table 10). Scenario 1 was also confirmed as the most likely one by model check analysis (Supplementary Figure 10A).

### Admixture Between ITAP and WCEUR Populations

The second question related to the gene flow between ITAP and WCEUR populations (Figure 6B) established Scenario 6 as the best one, due to the highest logistic regression values, assuming that an admixture event between ITAP-south and WCEUR genotypes flowed into ITAP-north-center



**FIGURE 5 |** Ancestry coefficient evaluated by R/TESS3 for the whole grapevine germplasm including 21 samples from North Africa (Morocco and Tunisia) named MAGH. The number of ancestral populations ( $K = 8$ ) was chosen after the evaluation of a cross-entropy criterion for each  $K$  (Supplementary Figure 9). (A) The values of the Q-matrix ( $K = 8$ ) were interpolated on a geographic map, using the centroid for each country studied, except for Russia for a better graphic display. Likewise, the samples from Bosnia and Herzegovina, Croatia, Montenegro, and Serbia were grouped together for a clearer visualization. (B) Bar plot showing the ancestry proportions contributed by  $K$  ancestral source populations by calculating the average value of each of the 8 membership coefficients (Q-value) for each of the seven groups defined by geographic origin (Table 1 and Supplementary Table 1).

germplasm (Supplementary Figure 9B). Scenario 6 showed the highest and significant Pp-value (81%), with a 95% CI ranging from 75 to 82%, 3 out of 9 statistics outlying from the observed values (Table 2 and Supplementary Table 9), 87% of performance evaluation, and Type I and Type II error rates <20% (Supplementary Table 10). The model check analysis confirmed the goodness-of-fit of Scenario 6 (Supplementary Figure 10B).

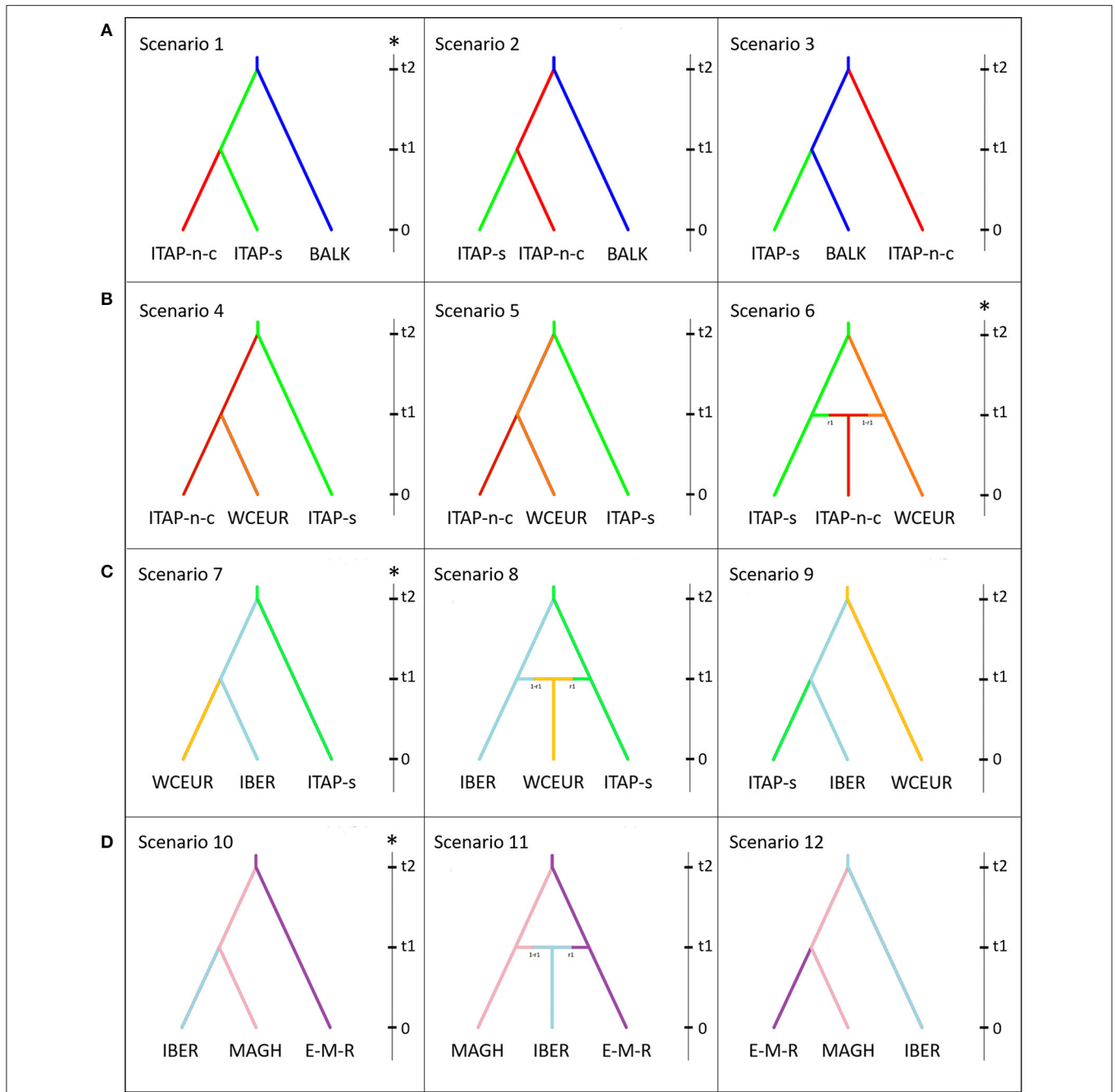
### The Gene Flow From ITAP-South to WCEUR Populations Through IBER Population

Among the scenarios hypothesized to evaluate the possibility of a gene flow from ITAP-south to WCEUR population through IBER population (Figure 6C), Scenario 7 showed the highest logistic regression values, assuming that grapevine

germplasm from ITAP-south was firstly introduced into IBER and then into WCEUR (Supplementary Figure 9C). Scenario 7 showed the highest Pp rate (88%, with 95% CI from 82 to 91%), the lowest number of outlying statistics (3 out of 9) (Table 2 and Supplementary Table 9), the highest performance evaluation value (92%), and Type I and Type II error rates <20% (Supplementary Table 10). Scenario 7 was confirmed as the most likely by model check analysis (Supplementary Figure 10C).

### The Gene Flow From EMCA-MFEAS-RUUK to IBER Populations Through MAGH Population

The last ABC test confirmed Scenario 10 as the most likely model (Figure 6D) taking into account the highest rate of



**FIGURE 6 |** Schematic representation of scenarios depicting the grapevine gene flow in the Mediterranean basin, based on ABC approach. Scenarios were depicted to validate or rebut the gene flow. **(A)** from Balkans to Northern Italy, **(B)** from ITAP-south to Western and Central Europe, **(C)** from ITAP-south to WCEUR through IBER, and **(D)** from Middle and Far East to IBER through Northern Africa. The vertical bar in each box indicates the timescale ( $t_2 > t_1 > t_0$ ;  $t_0$  = present). For each hypothesis, the most likely scenario (\*) was selected among the ones showing: (i) the highest values of logistic regression; (ii) the lowest number of outlying statistics; (iii) the highest value of posterior probability; (iv) the highest value of performance evaluation; (v) the mean values of Types I-II error lower than 20%, indicating an adequate power and sensitivity of ABC analysis (Chen et al., 2017). The samples named EMCA-MFEAS-RUUK (Table 1) was indicated as E-M-R, while ITAP-north-center and ITAP-south was named, ITAP-n-c and ITAP-s, respectively. ABC, approximate Bayesian computation; EMCA-MFEAS-RUUK; Eastern Mediterranean and Caucasus, Middle and Far East, Russia and Ukraine; ITAP-south, ITAP-Southern Italy; ITAP-north-center, ITAP-Northern, and Center Italy; WCEUR, Western and Central Europe.

logistic regression and assuming a gene flow from EMCA-MFEAS-RUUK to Northern Africa (MAGH) and then to the IBER population (Supplementary Figure 9D). Scenario 10

showed 75% Pp (with 95% CI from 71 to 79%), the lowest number of outlying statistics (4 out of 9), high value of performance evaluation (88%), and Type I and Type II error



**TABLE 2** | Summary of ABC results to define the most likely scenarios belonging to four hypotheses about the gene flow of grapevine genetic resources from East-to-West.

Hypothesis 1	Scenario 1	Scenario 2	Scenario 3
Pp (%)	<b>74</b>	51	43
95% CI	<b>93–95</b>	48–54	40–45
Number of outlying statistics			
<i>P</i> < 0.05	<b>1</b>	6	2
<i>P</i> < 0.01	<b>2</b>	0	3
Hypothesis 2	Scenario 4	Scenario 5	Scenario 6
Pp (%)	52	57	<b>81</b>
95% CI	46–53	55–64	<b>75–82</b>
Number of outlying statistics			
<i>P</i> < 0.05	0	2	<b>1</b>
<i>P</i> < 0.01	4	2	<b>2</b>
Hypothesis 3	Scenario 7	Scenario 8	Scenario 9
Pp (%)	<b>88</b>	57	38
95% CI	<b>82–91</b>	54–64	32–40
Number of outlying statistics			
<i>P</i> < 0.05	<b>1</b>	2	2
<i>P</i> < 0.01	<b>2</b>	2	2
Hypothesis 4	Scenario 10	Scenario 11	Scenario 12
Pp (%)	<b>75</b>	46	58
95% CI	<b>71–79</b>	42–48	57–60
Number of outlying statistics			
<i>P</i> < 0.05	<b>2</b>	3	3
<i>P</i> < 0.01	<b>2</b>	4	2

Scenarios are defined in **Figure 6**. Pp, posterior probability; CI, confidence interval. Scenarios with highest Pp-value and lowest number of outlying statistics within each hypothesis are in bold and italic.

rates <20% (**Table 2, Supplementary Tables 9, 10**). The PCA analysis performed in the model check analysis confirmed Scenario 10 as the scenario well-fitting the observed dataset (**Supplementary Figure 10D**).

## DISCUSSION

Significant efforts were made over the last decades to clarify the complex puzzle of grapevine germplasm evolution and spreading in Eurasia. To trace the history of a perennial crop, the evolution rate should be considered. In grapevine, it is expected to be slow due to the limited sexual reproduction and clonal propagation (Miller and Gross, 2011; Lacombe et al., 2012). Archaeological evidence suggested that grapevine domestication took place in the Near East between the Black and Caspian Sea, which has been confirmed by the genetic analysis (Arroyo-García et al., 2006; This et al., 2006; Myles et al., 2011; McGovern et al., 2017). Moreover, the grapevine domestication appeared dynamic in space and time, defining a continuum from exploited wild to cultivated populations, *via* incipient domesticated populations (Miller and Gross, 2011).

In the grapevine, the coexistence of wild populations and domesticated varieties is well-documented with a bidirectional gene flow (De Andrés et al., 2012; Riaz et al., 2018;

D’Onofrio, 2020; Maraš et al., 2020). The consequent generations overlapping due to hybridization events between wild and modern varieties implied stratification that cannot be resolved on a time scale, as revealed by pedigree analysis (Lacombe et al., 2012; D’Onofrio, 2020). Furthermore, the synonymies and homonymies in the cultivated germplasm and the difficulty in assigning the varieties to specific areas made the gene flow studies even more complex. Grapevine is propagated by cuttings; to propagate desirable cultivars indefinitely, these cuttings were introduced into different regions, following human migrations, and were renamed locally (This et al., 2006). In the last few decades, a massive clearing up of synonyms and homonyms and pedigree reconstruction by using molecular markers (SSR, SNP) became necessary. Moreover, the availability of this information on public databases provided advanced knowledge on grapevine varieties supporting their assignment to geographic areas, supplying a decisive input to gene flow studies.

A great deal of information from a large international cultivars panel was recently obtained, although fragmented due to the use of different genotype panels as well as by molecular marker sets (This et al., 2004; Cipriani et al., 2010; Myles et al., 2011; Cunha et al., 2016; Laucou et al., 2018; Riaz et al., 2018; De Lorenzis et al., 2019; D’Onofrio et al., 2021). The relationships within the grapevine germplasm were drawn worldwide, even though not completely. Here, we merged the grapevine genetic resources

from two Italian projects (AGER and VIGNETO), resulting in 384 genotyped varieties using the 18k SNP genotyping array. Varieties belonging to Eurasia, from the Caucasus to Western Europe (De Lorenzis et al., 2015, 2019; Laucou et al., 2018), were added for the final dataset of 1,038 unique SNP-profiles.

## History of the Italian Population and Distinctness From the Other Eurasian Populations

The origin of Italian viticulture dates to the Etruscans (around the 8<sup>th</sup> BC) and Greeks (7<sup>th</sup>-6<sup>th</sup> BC). The former domesticated wild grapevines in ITAP-north-center, while the Greeks introduced their varieties in ITAP-south. Differences between these areas occurred in winegrowing, the choice of varieties, and in the way to manage the vines (Buono and Vallariello, 2002), but these differences were not highlighted yet by genetic analysis. Furthermore, the Italian varieties did not appear distinguishable from those of the other European regions. The first research including a considerable panel of Italian varieties, genotyped at 34 SSR loci, was unable to cluster the germplasm based on their geographic origin (Cipriani et al., 2010). Later, 304 Italian varieties, included in a large germplasm collection from the Information System of the National Institute for Agronomical Research (INRA), were characterized by using another panel of 20 SSR loci, resulting admixed by a weak genetic structure based on geographic origin (Bacilieri et al., 2013). Recently, 783 Eurasian varieties belonging to four grapevine repositories were characterized with the 10k genome-wide SNPs and the 84 Italian cultivars clustered together with WCEUR genotypes, confirming the indiscernibility of the Italian germplasm (Laucou et al., 2018). Then, 140 varieties from ITAP-south characterized by the same 18k SNPs showed similarity with genotypes coming from both Eastern Mediterranean Sea and France, although the ITAP-south cluster appeared evident (De Lorenzis et al., 2019). D'Onofrio et al. (2021) highlighted that most of the Italian varieties, genotyped by SNPs, appeared structured in few main clusters with the first and second-degree relationship scattered from the South to North of Italy.

Here, for the first time, the cluster analysis on 384 Italian genotypes allowed us to distinguish the Northern and Southern genotypes, appearing clearly separated into two large clusters by using both UPGMA algorithm (Figure 1A) and PCoA (Supplementary Figure 1). The observed distinctiveness was better detailed by using a multiple Bayesian approach, DAPC (Figures 1B,C and Supplementary Table 1). A significant  $K = 3$  was evidenced with a group (Pool 2) including South-Italian varieties and two others overlapped, the first mainly counting North-Italian varieties (Pool 1), and the second (Pool 3) including varieties from all Italian areas (Figure 1B). In each pool, some ancestors can be recognized, such as Garganega/Greccanico dorato in Pool 1, Bombino and Visparola in Pool 3 (D'Onofrio et al., 2021) and Mantonico bianco and Sangiovese in Pool 2 (De Lorenzis et al., 2019; D'Onofrio et al., 2021) (Supplementary Table 1). These last could represent “bridge varieties” linking ITAP-south and ITAP-north-center populations.

The complex origin of ITAP-south germplasm was confirmed; indeed the pairwise fixation index,  $F_{st}$  was the same in both couples ITAP-south/BALK and ITAP-south/ITAP-north-center (Supplementary Table 5 and Figure 3A). The higher proximity of ITAP-south to BALK than to ITAP-north-center germplasm was inferred by other analyses, such as MDS with IBD, NJ tree, and TreeMix (Figures 3B,D, 4). Otherwise, ITAP-north-center varieties showed a higher proximity to WCEUR germplasm (Figures 3A–C, 4C). The  $F_{st}$  in the couple ITAP-north-center/WCEUR was the lowest among all pairwise comparisons (Supplementary Table 5). Moreover, one major migration event from Southern to Northern Italy, inferred with TreeMix analysis (Figure 4C), is supported by a recent pedigree study (D'Onofrio et al., 2021) and the possible “bridge” role of the varieties included in Pool 3 (DAPC) (Figures 1B,C and Supplementary Table 1). In addition, Scenario 1 (ABC analysis) agrees with this hypothesis, as the best supported scenario for a mixed origin of ITAP-north-center germplasm between ITAP-south and WCEUR genotypes (Figure 6B). The ITAP-north-center and ITAP-south genotypes were distinguished, also when compared to the Eurasian dataset. Indeed, the DAPC on the whole SNP dataset (1,038 genotypes) split the ITAP germplasm in two groups (Figures 2B,C and Supplementary Table 1). Following the posterior membership probabilities (>70%), Pools 6 and 7 included mainly varieties from WCEUR and ITAP-north-center, respectively (Supplementary Table 1). Once again, ITAP-north and ITAP-south varieties were distinguishable, while ITAP-center germplasm was split. Small groups of Italian varieties were pinpointed in other pools, e.g., some Sardinian cultivars, historically raked from Spain to Sardinia (De Mattia et al., 2009), grouped in Pool 8, representative of the IBER germplasm. Some muscat flavored varieties grouped in Pool 5, where the offspring of the well-recognized major ancestor Moscato bianco/Muscat à petits grains blanc were placed (Cipriani et al., 2010; Ruffa et al., 2016). Finally, the geographic location can account for some cultivars from the Adriatic coast grouped into Pool 2, including BALK germplasm.

A slightly more complex picture was defined by ancestry coefficient analysis (R/TESS3) on the whole dataset, where ITAP-south germplasm resulted as the main contributor to  $K_5$ , while ITAP-north-center cultivars showed the highest values of membership in two pools,  $K_4$  and  $K_7$ , with  $K_7$  appearing as a mixed pool (Figure 5 and Supplementary Table 8).

## Genetic Structure of the Eurasian Germplasm

The genetic distances among the six Eurasian populations (Figure 3A) indicated the highest divergence between EMCA-MFEAS-RUUK and WCEUR populations, which was also confirmed by the highest  $F_{st}$  index among all pairwise comparisons (Supplementary Table 5). This result was expected given the well-known gene flow from Eastern to Western Eurasia. Our results appeared in agreement with an evoked divergence among the grapevine genetic structures in different European grapevine-growing regions, firstly reported by Sefc

et al. (2000). Bayesian computations helped us depict a more detailed reconstruction of this history.

The DAPC performed on the whole dataset highlighted nine pools, attributable to the Eurasian populations (**Figures 2B,C**). Seven out of nine pools appeared correlated with the main geographic memberships (**Supplementary Table 1**). Indeed, Pools 6 and 7 included nearly all Italian varieties, Pool 8 included the highest number of varieties from IBER, Pools 3 and 9 represented the EMCA-MFEAS-RUUK germplasm; noticeably, the Georgian varieties were grouped in Pool 9, while Pools 1 and 4 included many varieties from WCEUR population. Finally, Pool 2 comprised many varieties of mixed geographic origin, two main parents, Gouais blanc/Heunischweiss and Blank Blauer/Vulpea, ascribed to WCEUR, but also a predominance of BALK varieties. The geographic membership of Pool 2 varieties is still questioned (Bowers et al., 1999; Crespan et al., 2020), and our results suggested their potential “bridge” role. Interestingly, Pool 5 resulted mixed between IBER and WCEUR germplasm putatively explaining the migration event from IBER to WCEUR population, highlighted by TreeMix analysis (**Figure 4C**) and supported in ABC results on Scenario 7 (**Figure 6C**). This appears to be a new finding in contrast with previous reports that hypothesized a migration of some cultivars from WCEUR to IBER population after a secondary domestication event in IBER (Arroyo-García et al., 2006).

The ancestry coefficient analysis highlighted one less genetic pool than DAPC ( $K = 8$ ) of which seven pools showed the highest ancestry membership values in agreement with the geographic origin. Similarly, Pools from 1 to 5 were assigned to IBER, EMCA-MFEAS-RUUK, Georgian, ITAP-north-center, and ITAP-south populations, respectively (**Figure 5B** and **Supplementary Table 8**). Pool 6 was mainly shared by the BALK varieties, deserving additional comments: three varieties from Cyprus harbor a very low ancestry and should have been ascribed to EMCA-MFEAS-RUUK. Notably, the Greek varieties showed multiple memberships, with similar ancestry values for  $K2$  (including EMCA-MFEAS-RUUK germplasm) and  $K6$  pools. Therefore, Greece appears to be the main bridge between Near Eastern germplasm and the Mediterranean Basin (ITAP), supporting the theory that grapevines were spread from the Greek shores to the Southern and then Northern Italian ones (Buono and Vallariello, 2002). The  $K7$  pool resulted largely admixed, in which Swiss varieties showed the highest ancestry value, followed by ITAP-north-center. The  $K8$  pool encompassed the WCEUR germplasm. Unexpectedly, the Portuguese varieties showed a greatest membership to  $K8$  (**Supplementary Table 8**), higher than  $K1$  (IBER pool), appearing to be the most plausible candidate group for the genetic exchange between IBER and WCEUR, highlighted by TreeMix computations and DAPC analysis. This hypothesis is also supported by ancestry coefficient analysis, where Portugal is painted with two colors (**Figure 5A**).

In summary, IBER germplasm appeared distinguished from the others indicating a complex history supported by genetic relationship (**Figure 2** and **Supplementary Figure 4**), IBD, and NJ analyses (**Figures 3B–D**). The genetic structure analysis stated that IBER varieties included two groups, one related to WCEUR group and the other to Northern Africa and

Eastern varieties (Bacilieri et al., 2013). Through structure analysis, ten IBER varieties (**Supplementary Table 1**) showed a larger ancestry membership to Pool 3 (EMCA-MFEAS-RUUK) than pool 8 (IBER). They could have been derived from the migration event from EMCA-MFEAS-RUUK to IBER, highlighted by TreeMix analysis (**Figure 4C**) and mediated by MAGH, as hypothesized in Scenario 10 by ABC analysis (**Figure 6D**). The groups defined by DAPC and the ancestry coefficient analyses sustained a major migration event from IBER to WCEUR population (TreeMix), explained by Portuguese varieties falling in Pool 4 (WCEUR) (**Supplementary Table 1**), Pool 5 (WCEUR and IBER) (**Supplementary Table 4**) and  $K8$  (**Supplementary Table 8**), respectively. Moreover, the Scenario 7 from ABC analysis well-supported a gene flow from ITAP-south to IBER and then to WCEUR population (**Figure 6C**), as hypothesized by Buono and Vallariello (2002).

### LD-Values Differed in the Six Geographic Groups

Standard  $r^2$ -value used to perform LD measure is a parameter suffering from bias when the individuals are dependent, or the analyzed populations evolved differently. In such cases, long-range values for LD are obtained as admixed populations (Mangin et al., 2012). Therefore, the distinctiveness of the evolution of six grape populations in their different geographic contexts, leading to differences in the allele frequencies affecting  $r^2$ -values appeared indubitable. Accordingly, shorter LD-values can account for a higher number of meiotic events and recombination within the population, leading to limited relationships. By contrast, high levels of admixture, close relationships, and low recombination rates, lead to longer LD-values (Mangin et al., 2012).

Providing that these observations are not significant on a time scale; according to LD trend, EMCA-MFEAS-RUUK population showed a more natural and faster evolution, followed by BALK and ITAP germplasm, while IBER and WCEUR varieties appeared more static and closely related (**Figure 3C**). The admixture within WCEUR and IBER populations, and the gene flow from WCEUR to IBER germplasm, were already established by pedigree analyses (Bowers et al., 1999; Boursiquot et al., 2009; Diaz Losada et al., 2010; Zinelabidine et al., 2015). Moreover, archaeological research in France highlighted that vines remain closely related to the Western European cultivars used nowadays for winemaking. The varieties cultivated by the Romans in ancient times were the same as those grown today or shown to be parent-offspring related with present varieties (Ramos-Madrigril et al., 2019).

### The East-to-West Grapevine Gene Flow

Grapevine domestication seems to be related to winemaking, although which process predated the other remains questionable (Terral et al., 2010). The pivotal role of Greece in the viticulture history is well-known. Indeed, the Greeks achieved mastery in grapevine cultivation and winemaking, elevating both to a cultural phenomenon. There is no reliable information on the introduction of winemaking techniques in Greece. From Greece, the culture of wine reached Western Europe through



Southern Italy and then France and Spain (Buono and Vallariello, 2002; Scienza, 2004). This migration flow from the East to West also involved grapevine cultivars, recently documented by genetic analysis (Myles et al., 2011; De Lorenzis et al., 2019; D'Onofrio et al., 2021). Our results confirmed this East-to-West gene flow. The population pairwise *Fst* estimation assigned the lowest genetic distance to ITAP-north-center and WCEUR pair, while the highest was found between WCEUR and EMCA-MFEAS-RUUK (Supplementary Table 5). The scaling of NJ-tree branches related to WCEUR population reflected a more recent development (Figure 3A). What has happened in between has to be analyzed. The complex role played by Italian germplasm evoked in this work, suggested a key breakthrough to understanding the grapevine germplasm evolution in Eurasia. Genetic data analysis (Figures 3A, 4) confirmed gene flow events from EMCA-MFEAS-RUUK to BALK and then to ITAP-south population. The novelty of the findings is the relationship between ITAP-north-center and WCEUR populations (Figures 3A–C, 4). The ABC analysis suggested that the admixture of ITAP-north-center germplasm could be derived by the gene flow from both ITAP-south and WCEUR varieties (Figure 6B, Scenario 6).

The gene flow from ITAP-south to WCEUR throughout IBER population assigned an additional role to the *Magna Graecia* (Figures 4C, 6C, Scenario 7). The complex structure of IBER population is also affected by the additional gene flow from EMCA-MFEAS-RUUK through the MAGH population, as highlighted by gene flow, ancestry evaluation, and ABC analysis through Scenario 10 (Figures 4, 5, 6D). As expected, the WCEUR population was the last member of this chain, showing the highest distance from the EMCA-MFEAS-RUUK group. In turn, WCEUR varieties played a role in gene flow, receiving a contribution from IBER, and contributing to the ITAP-north-center germplasm (Figures 4, 6, Scenario 6).

After the first grapevine domestication event in the Caucasus, cultivated vines spread southwards to Anatolia and Egypt 5,000 years ago, as well as around the Mediterranean Basin, following the main civilizations (McGovern, 2003; This et al., 2006). In this scenario, the key role of Italian germplasm in the grape genetic differentiation (Grassi et al., 2003; Sunseri et al., 2018; De Lorenzis et al., 2019)

was confirmed. We provide a sophisticated genetic model that is useful to ultimately attribute the role of frontier between the Western and Eastern Eurasia to the Italian and mainly to the *Magna Graecia* germplasm.

Our finding, reconciling genetic and archaeological data for one of the most cultivated and fascinating crops in the world, has made a key contribution to defining the genetic relationships among grapevine populations distributed in a wide geographic area ranging from the Caucasus to the Iberian Peninsula.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

## AUTHOR CONTRIBUTIONS

FM, GDL, MC, and FS: conceptualization. FM and GDL: methodology, formal analysis, and writing—original draft. AM, MZ, CM, PR, LB, CD'O, MB, CB, LP, and VN: investigation. CM, PR, LB, CD'O, MB, CB, LP, and VN: resources. FM, MC, and FS: supervision. CM, PR, LB, CD'O, MB, CB, LP, FS, and VN: funding acquisition. All authors contributed to the writing—review and editing, article, and approved the submitted version.

## FUNDING

This work was supported by an Ager grant n 2010–2104: An Italian *Vitis* database with multidisciplinary approach, for exploitation and valorization of the regional genotypes and the CREA grant, Viticultural Characterization of the main Italian Grape Varieties and their Terroir (VIGNETO) project.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.692661/full#supplementary-material>

## REFERENCES

- Arroyo-García, R., Ruiz-García, L., Bolling, L., Ocete, R., López, M., a, Arnold, C., et al. (2006). Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *sativa*) based on chloroplast DNA polymorphisms. *Mol. Ecol.* 15, 3707–3714. doi: 10.1111/j.1365-294X.2006.03049.x
- Bacilieri, R., Lacombe, T., Le Cunff, L., Di Vecchi-Staraz, M., Laucou, V., Genna, B., et al. (2013). Genetic structure in cultivated grapevines is linked to geography and human selection. *BMC Plant Biol.* 13:25. doi: 10.1186/1471-2229-13-25
- Boursiquot, J.-M., Lacombe, T., Laucou, V., Julliard, S., Perrin, F.-X., Lanier, N., et al. (2009). Parentage of Merlot and related winegrape cultivars of southwestern France: discovery of the missing link. *Aust. J. Grape Wine Res.* 15, 144–155. doi: 10.1111/j.1755-0238.2008.00041.x
- Bowers, J., Boursiquot, J.-M., This, P., Chu, K., Johansson, H., and Meredith, C. (1999). Historical genetics: the parentage of chardonnay, gamay, and other wine grapes of Northeastern France. *Science* 285, 1562–1565. doi: 10.1126/science.285.5433.1562
- Buono, R., and Vallariello, G. (2002). Introduzione e diffusione della vite (*Vitis vinifera* L.) in Italia. *Delpinoa* 44, 39–51.
- Caye, K., Deist, T. M., Martins, H., Michel, O., and François, O. (2016). TESS3: fast inference of spatial population structure and genome scans for selection. *Mol. Ecol. Resour.* 16, 540–548. doi: 10.1111/1755-0998.12471
- Chen, M., Fontaine, M. C., Ben Chehida, Y., Zheng, J., Labbé, F., Mei, Z., et al. (2017). Genetic footprint of population fragmentation and contemporary collapse in a freshwater cetacean. *Sci. Rep.* 7:14449. doi: 10.1038/s41598-017-14812-z
- Cipriani, G., Spadotto, A., Jurman, I., Di Gasparo, G., Crespan, M., Meneghetti, S., et al. (2010). The SSR-based molecular profile of 1005 grapevine (*Vitis vinifera*

- L.) accessions uncovers new synonymy and parentages, and reveals a large admixture amongst varieties of different geographic origin. *Theor. Appl. Genet.* 121, 1569–1585. doi: 10.1007/s00122-010-1411-9
- Cornuet, J.-M., Pudlo, P., Veysier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., et al. (2014). DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* 30, 1187–1189. doi: 10.1093/bioinformatics/btt763
- Crespan, M., Migliaro, D., Larger, S., Pindo, M., Petrusi, C., Stocco, M., et al. (2020). Unraveling the genetic origin of 'Glera', 'Ribolla Gialla' and other autochthonous grapevine varieties from Friuli Venezia Giulia (northeastern Italy). *Sci. Rep.* 10:7206. doi: 10.1038/s41598-020-64061-w
- Cunha, J., Ibáñez, J., Teixeira-Santos, M., Brazão, J., Feveiro, P., Martínez-Zapater, J. M., et al. (2016). Characterisation of the Portuguese grapevine germplasm with 48 single-nucleotide polymorphisms. *Aust. J. Grape Wine Res.* 22, 504–516. doi: 10.1111/ajgw.12225
- De Andrés, M. T., Benito, A., Pérez-Rivera, G., Ocete, R., Lopez, M., Gaforio, L., et al. (2012). Genetic diversity of wild grapevine populations in Spain and their genetic relationships with cultivated grapevines. *Mol. Ecol.* 21, 800–816. doi: 10.1111/j.1365-294X.2011.05395.x
- 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 Biol.* 15:154. doi: 10.1186/s12870-015-0510-9
- De Lorenzis, G., Mercati, F., Bergamini, C., Cardone, M. F., Lupini, A., Mauceri, A., et al. (2019). SNP genotyping elucidates the genetic diversity of *Magna Graecia* grapevine germplasm and its historical origin and dissemination. *BMC Plant Biol.* 19:7. doi: 10.1186/s12870-018-1576-y
- De Mattia, F., Lovicu, G., Tardaguila, J., Grassi, F., Imazio, S., Scienza, A., et al. (2009). Genetic relationships between Sardinian and Spanish viticulture: the case of 'Cannonau' and 'Garnacha'. *J. Hortic. Sci. Biotechnol.* 84, 65–71. doi: 10.1080/14620316.2009.11512481
- Diaz Losada, E., Tato Salgado, A., Ramos-Cabrera, A. M., Rio Segade, S., Cortes Dieguez, S., and Pereira-Lorenzo, S. (2010). Twenty microsatellites (SSRs) reveal two main origins of variability in grapevine cultivars from Northwestern Spain. *Vitis J. Grapevine Res.* 49, 55–62.
- D'Onofrio, C. (2020). Introgression among cultivated and wild grapevine in tuscany. *Front. Plant Sci.* 11:202. doi: 10.3389/fpls.2020.00202
- D'Onofrio, C., Tumino, G., Gardiman, M., Crespan, M., Schneider, A., Bignami, C., et al. (2021). Parentage atlas of widely and locally cultivated Italian grapevine varieties as inferred from SNP genotyping. *Front. Plant Sci.* 11:605934. doi: 10.3389/fpls.2020.605934
- Emanuelli, F., Lorenzi, S., Grzeskowiak, L., Catalano, V., Stefanini, M., Troggo, M., et al. (2013). Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. *BMC Plant Biol.* 13:39. doi: 10.1186/1471-2229-13-39
- Fitak, R. (2019). *OptM: An R Package to Optimize the Number of Migration Edges Using Threshold Models*. Available online at: <https://cran.r-project.org/web/packages/OptM/index.html>
- Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., and François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics* 196, 973–983. doi: 10.1534/genetics.113.160572
- Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* 5, 184–186. doi: 10.1111/j.1471-8286.2004.00828.x
- Grassi, F., Labra, M., Imazio, S., Spada, A., Sgorbati, S., Scienza, A., et al. (2003). Evidence of a secondary grapevine domestication centre detected by SSR analysis. *Theor. Appl. Genet.* 107, 1315–1320. doi: 10.1007/s00122-003-1321-1
- Hill, W. G., and Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* 38, 226–231. doi: 10.1007/BF01245622
- Jombart, T. (2008). ADEGENET: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405. doi: 10.1093/bioinformatics/btn129
- Kamvar, Z. N., Brooks, J. C., and Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Genet.* 6:208. doi: 10.3389/fgene.2015.00208
- Lacombe, T., Boursiquot, J.-M., Laucou, V., Di Vecchi-Staraz, M., Pérois, J.-P., and This, P. (2012). Large-scale parentage analysis in an extended set of grapevine cultivars (*Vitis vinifera* L.). *Theor. Appl. Genet.* 126, 401–414. doi: 10.1007/s00122-012-1988-2
- Laucou, V., Lacombe, T., Dechesne, F., Siret, R., Bruno, J.-P., Dessup, M., et al. (2011). High throughput analysis of grape genetic diversity as a tool for germplasm collection management. *Theor. Appl. Genet.* 122, 1233–1245. doi: 10.1007/s00122-010-1527-y
- Laucou, V., Launay, A., Bacilieri, R., Lacombe, T., Adam-Blondon, A.-F., Bérard, A., et al. (2018). Extended diversity analysis of cultivated grapevine *Vitis vinifera* with 10K genome-wide SNPs. *PLoS ONE* 13:e0192540. doi: 10.1371/journal.pone.0192540
- Levadoux, L. (1956). Les populations sauvages et cultivées de *Vitis vinifera* L. *Ann. Lamélor. Plant.* 1, 59–118.
- Mangin, B., Siberchicot, A., Nicolas, S., Doligez, A., This, P., and Cierco-Ayrolles, C. (2012). Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity (Edinb)*. 108, 285–291. doi: 10.1038/hdy.2011.73
- Maraš, V., Tello, J., Gazivoda, A., Mugoša, M., Perišić, M., Raičević, J., et al. (2020). Population genetic analysis in old Montenegrin vineyards reveals ancient ways currently active to generate diversity in *Vitis vinifera*. *Sci. Rep.* 10:15000. doi: 10.1038/s41598-020-71918-7
- Marrano, A., Birolo, G., Prazzoli, M. L., Lorenzi, S., Valle, G., and Grando, M. S. (2017). SNP-discovery by RAD-sequencing in a germplasm collection of wild and cultivated grapevines (*V. vinifera* L.). *PLoS ONE* 12:e0170655. doi: 10.1371/journal.pone.0170655
- Marvelli, S., De' Siena, S., Rizzoli, E., and Marchesini, M. (2013). The origin of grapevine cultivation in Italy: the archaeobotanical evidence. *Ann. Bot.* 3, 155–163. doi: 10.4462/annbotrm-10326
- McGovern, P., Jalabadze, M., Batiuk, S., Callahan, M. P., Smith, K. E., Hall, G. R., et al. (2017). Early Neolithic wine of Georgia in the South Caucasus. *Proc. Natl. Acad. Sci. U.S.A.* 114, 10309–10318. doi: 10.1073/pnas.1714728114
- McGovern, P., Voigt, M., Glusker, D., and Exner, L. (1986). Neolithic resinated wine. *Nature* 381, 480–481. doi: 10.1038/381480a0
- McGovern, P. E. (2003). *Ancient Wine*. Princeton, NJ: Princeton University Press.
- Mercati, F., De Lorenzis, G., Brancadoro, L., Lupini, A., Abenavoli, M. R., Barbagallo, M. G., et al. (2016). High-throughput 18K SNP array to assess genetic variability of the main grapevine cultivars from Sicily. *Tree Genet. Genomes* 12:59. doi: 10.1007/s11295-016-1021-z
- Miller, A. J., and Gross, B. L. (2011). From forest to field: perennial fruit crop domestication. *Am. J. Bot.* 98, 1389–1414. doi: 10.3732/ajb.1000522
- Myles, S., Boyko, A. R., Owens, C. L., Brown, P. J., Grassi, F., Aradhya, M. K., et al. (2011). Genetic structure and domestication history of the grape. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3530–3535. doi: 10.1073/pnas.1009363108
- Myles, S., Chia, J.-M., Hurwitz, B., Simon, C., Zhong, G. Y., Buckler, E., et al. (2010). Rapid genomic characterization of the genus *Vitis*. *PLoS ONE* 5:e8219. doi: 10.1371/journal.pone.0008219
- Neath, A. A., and Cavanaugh, J. E. (2012). The Bayesian information criterion: background, derivation, and applications. *Wiley Interdiscip. Rev. Comput. Stat.* 4, 199–203. doi: 10.1002/wics.199
- Nei, M. (1972). Genetic distance between populations. *Am. Nat.* 106, 283–292. doi: 10.1086/282771
- Orozco-terWengel, P., Barbato, M., Nicolazzi, E., Biscarini, F., Milanese, M., Davies, W., et al. (2015). Revisiting demographic processes in cattle with genome-wide population genetic analysis. *Front. Genet.* 6:191. doi: 10.3389/fgene.2015.00191
- Pickrell, J. K., and Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8:e1002967. doi: 10.1371/journal.pgen.1002967
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Raj, A., Stephens, M., and Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197, 573–589. doi: 10.1534/genetics.114.164350
- Ramos-Madrugal, J., Runge, A. K. W., Boubry, L., Lacombe, T., Samaniego Castruita, J. A., Adam-Blondon, A.-F., et al. (2019). Palaeogenomic insights into the origins of French grapevine diversity. *Nat. Plants* 5, 595–603. doi: 10.1038/s41477-019-0437-5

- Riaz, S., De Lorenzis, G., Velasco, D., Koehmstedt, A., Maghradze, D., Bobokashvili, Z., et al. (2018). Genetic diversity analysis of cultivated and wild grapevine (*Vitis vinifera* L.) accessions around the Mediterranean basin and Central Asia. *BMC Plant Biol.* 18:137. doi: 10.1186/s12870-018-1351-0
- Ruffa, P., Raimondi, S., Boccacci, P., Abbà, S., and Schneider, A. (2016). The key role of “Moscato bianco” and “Malvasia aromatica di Parma” in the parentage of traditional aromatic grape varieties. *Tree Genet. Genomes* 12:50. doi: 10.1007/s11295-016-1006-y
- Scienza, A. (2004). “Magna Grecia e Sicilia,” in *La vite e l'uomo dal rompicapo delle origini al salvataggio delle reliquie*, eds F. Del Zan, O. Failla, and A. Scienza (Trieste: Editoriale Lloyd) 853–942.
- Sefc, K. M., Lopes, M. S., Lefort, F., Botta, R., Roubelakis-Angelakis, K. A., Ibáñez, J., et al. (2000). Microsatellite variability in grapevine cultivars from different European regions and evaluation of assignment testing to assess the geographic origin of cultivars. *Theor. Appl. Genet.* 100, 498–505. doi: 10.1007/s001220050065
- Sunseri, F., Lupini, A., Mauceri, A., De Lorenzis, G., Araniti, F., Brancadoro, L., et al. (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 Res.* 24, 345–359. doi: 10.1111/ajgw.12339
- Terral, J. F., Tabard, E., Bouby, L., Ivorra, S., Pastor, T., Figueiral, I., et al. (2010). Evolution and history of grapevine (*Vitis vinifera*) under domestication: new morphometric perspectives to understand seed domestication syndrome and reveal origins of ancient European cultivars. *Ann. Bot.* 105, 443–455. doi: 10.1093/aob/mcp298
- Thachuk, C., Crossa, J., Franco, J., Dreisigacker, S., Warburton, M., and Davenport, G. F. (2009). Core Hunter: an algorithm for sampling genetic resources based on multiple genetic measures. *BMC Bioinformatics* 10:243. doi: 10.1186/1471-2105-10-243
- This, P., Jung, A., Boccacci, P., Borrego, J., Botta, R., Costantini, L., et al. (2004). Development of a standard set of microsatellite reference alleles for identification of grape cultivars. *Theor. Appl. Genet.* 109, 1448–1458. doi: 10.1007/s00122-004-1760-3
- This, P., Lacombe, T., and Thomas, M. R. (2006). Historical origins and genetic diversity of wine grapes. *Trends Genet.* 22, 511–519. doi: 10.1016/j.tig.2006.07.008
- Wright, T. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution (N. Y.)* 19, 395–420. doi: 10.1111/j.1558-5646.1965.tb01731.x
- Zecca, G., Labra, M., and Grassi, F. (2020). Untangling the evolution of American wild grapes: admixed species and how to find them. *Front. Plant Sci.* 10:1814. doi: 10.3389/fpls.2019.01814
- Zhao, H., Nettleton, D., Soller, M., and Dekkers, J. C. M. (2005). Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. *Genet. Res.* 86, 77–87. doi: 10.1017/S001667230500769X
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28, 3326–3328. doi: 10.1093/bioinformatics/bts606
- Zinelabidine, L. H., Cunha, J., Eiras-Dias, J. E., Cabello, F., Martínez-Zapater, J. M., and Ibáñez, J. (2015). Pedigree analysis of the Spanish grapevine cultivar “Hebén.” *Vitis J. Grapevine Res.* 54, 81–86.
- Zohary, D., and Hopf, M. (2000). *Domestication of Plants in the Old World: The Origin and Spread of Cultivated Plants in West Asia, Europe, and the Nile Valley*. Oxford: Oxford University Press.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Mercati, De Lorenzis, Mauceri, Zerbo, Brancadoro, D'Onofrio, Morcia, Barbagallo, Bignami, Gardiman, de Palma, Ruffa, Novello, Crespan and Sunseri. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.