



OPEN ACCESS

EDITED BY

Cinzia Montemurro,
University of Bari Aldo Moro, Italy

REVIEWED BY

Fabrizio Carbone,
Council for Agricultural Research and
Economics (CREA), Italy
Francesca Taranto,
CNR-IBBR Institute of Bioscience and
BioResources, Italy
Clizia Villano,
University of Naples Federico II, Italy

*CORRESPONDENCE

Filippos Aravanopoulos,
✉ aravanop@auth.gr

RECEIVED 21 September 2023

ACCEPTED 20 November 2023

PUBLISHED 04 December 2023

CITATION

Tourvas N, Ganopoulos I, Koubouris G,
Kostelenos G, Manthos I, Bazakos C,
Stournaras V, Molassiotis A and
Aravanopoulos F (2023), Wild and
cultivated olive tree genetic diversity in
Greece: a diverse resource in danger
of erosion.

Front. Genet. 14:1298565.

doi: 10.3389/fgene.2023.1298565

COPYRIGHT

© 2023 Tourvas, Ganopoulos, Koubouris,
Kostelenos, Manthos, Bazakos,
Stournaras, Molassiotis and
Aravanopoulos. 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.

Wild and cultivated olive tree genetic diversity in Greece: a diverse resource in danger of erosion

Nikolaos Tourvas¹, Ioannis Ganopoulos², Georgios Koubouris³,
George Kostelenos⁴, Ioannis Manthos⁵, Christos Bazakos^{2,6},
Vasileios Stournaras⁷, Athanassios Molassiotis⁸ and
Filippos Aravanopoulos^{1*}

¹Laboratory of Forest Genetics, Faculty of Agriculture, Forestry and Natural Environment, Aristotle University of Thessaloniki, Thessaloniki, Greece, ²Institute of Plant Breeding and Genetic Resources, Hellenic Agricultural Organization (ELGO) DIMITRA, Thessaloniki-Thermi, Greece, ³Institute of Olive Tree, Subtropical Crops and Viticulture, Hellenic Agricultural Organization (ELGO) DIMITRA, Chania, Greece, ⁴Kostelenos Olive Nurseries, Poros-Trizinias, Greece, ⁵Department of Nut Trees, Institute of Plant Breeding and Genetic Resources, Hellenic Agricultural Organization (ELGO) DIMITRA, Neo Krikello-Lamia, Greece, ⁶Department of Comparative Development and Genetics, Max Planck Institute for Plant Breeding Research, Cologne, Germany, ⁷Department of Olive and Horticultural Crops, Institute of Olive Tree, Subtropical Crops and Viticulture, Hellenic Agricultural Organization (ELGO) DIMITRA, Kalamata, Greece, ⁸Laboratory of Pomology, Department of Horticulture, Aristotle University of Thessaloniki, Thessaloniki-Thermi, Greece

The genetic relationships between Greek wild olive tree populations and cultivars were investigated. A total of 219 wild genotypes and 67 cultivar genotypes were analyzed by employing 10 SSR markers. Data evidenced that the wild populations exhibited high levels of genetic diversity and exclusively host 40% of the total number of alleles detected. Inbreeding was observed within populations, probably as a consequence of their fragmented spatial distribution. The genetic differentiation between cultivars and wild individuals, as well as within wild populations, was low. Nevertheless, three gene pools of wild trees were detected, corresponding to the geographical areas of Northeastern Greece, Peloponnese-Crete and Epirus. Most cultivars clustered in a separate group, while the rest of them formed a heterogenous group with membership coefficients akin to the three wild olive clusters. Regarding the history of olive cultivation in Greece, bidirectional gene flow was detected between populations of Peloponnese-Crete and the gene pool that composes some of Greece's most important cultivars, such as "Koroneiki" and "Mastoidis", which is inferred as an indication of a minor domestication event in the area. A strategy for the protection of Greek-oriented olive genetic resources is proposed, along with suggestions for the utilization of the genetically diverse wild resources with regard to the introgression of traits of agronomical interest to cultivars.

KEYWORDS

crop evolution, cultivars, genetic diversity, genetic pollution, gene flow, microsatellites, *Olea europaea*, wild crop relatives

1 Introduction

The olive tree (*Olea europaea* subsp. *europaea*) is recognised as an emblematic plant species of the Mediterranean basin. It is found in two different types, known since antiquity as reported by Theophrastus (1916), wild (*O. europaea* subsp. *europaea* var. *sylvestris*) and cultivated (*O. europaea* subsp. *europaea* var. *europaea*) (Green, 2002). Both forms (wild and cultivated) are interfertile (Besnard et al., 2000) with 23 pairs ($2n = 2x = 46$) of chromosomes (Contento et al., 2002) and excellent grafting compatibility (Mulas and Deidda, 1998). The reproduction of wild forms takes place sexually, assisted by wind, as well as birds that contribute to seed dispersion (Alcántara and Rey, 2003). The presence-absence of wild olive trees is an excellent biomarker of the Mediterranean Floristic Region (Rubio et al., 2002). A plethora of cultivated accessions exist and are conserved in banks worldwide (Bartolini and Cerreti, 2008) and at the country level (Barranco and Rallo, 2000; Khadari et al., 2003; Gemas et al., 2004; Bracci et al., 2009; Haouane et al., 2011; Yoruk and Taskin, 2014; Koubouris et al., 2019). Based on the consensus from archeological, paleobotanical and genetic data, these cultivars originate almost entirely from ancestral populations of Mediterranean wild olive trees, also known as oleasters (Kaniewski et al., 2012; Besnard et al., 2018; Julca et al., 2023). Nowadays, oleasters are located in humid and subhumid thermo-Mediterranean areas where no frost is observed (Martínez and Gandullo, 1987; Carrión et al., 2010).

It is known from multiple studies that wild olive trees of the Mediterranean region are genetically distinguished into the Western (WW) and Eastern Mediterranean (WE) groups (Breton et al., 2006; Belaj et al., 2007; Diez et al., 2015). The diversity of cultivated genotypes is better described by the formation of three groups - Western (Q1), Central (Balkan and Italian peninsula - Q2), and Eastern Mediterranean (Q3) (Haouane et al., 2011; Diez et al., 2015; Besnard et al., 2018). Additionally, Q2 exhibits the highest chloroplast diversity compared to Q1 and Q3 (Diez et al., 2015), with the Aegean region as a hotspot. Two hypotheses have been proposed regarding the origin of the cultivated olive tree. The first describes a unique domestication event that occurred in the Eastern Mediterranean group (Q3) (Besnard et al., 2013a; 2013b; Khadari and El Bakkali, 2018). The main evidence in support of this hypothesis is the fact that 90% of the cultivated genotypes carry the Eastern Mediterranean chloroplast haplotype E1.1 (Besnard et al., 2013b). Nuclear DNA markers have confirmed the close genetic relationship of modern cultivars with the Eastern Mediterranean wild olive gene pool (Lumaret et al., 2004; Breton et al., 2006). However, other studies based on the results of nuclear microsatellite marker studies support the existence of two or more domestication events in the eastern (Q3) and central Mediterranean (Q2) groups (Breton et al., 2009; Diez et al., 2015). High-throughput sequencing analyzes (Gros-Balthazard et al., 2019; Julca et al., 2020) tend to support the existence of one major domestication event, but point to the existence of additional events of secondary magnitude. In contrast, Jiménez-Ruiz et al. (2020) analysed the transposable elements of cultivars and wild individuals and supported the presence of two independent domestication events. Recently, a more in-depth analysis of the olive genome, including copy number variations, was performed by Bazakos et al. (2023) and uncovered several candidate genes involved in selective sweeps.

Furthermore, a GWAS analysis for morphological and agronomic traits allowed the detection of significant associations to support the next steps of olive tree breeding (Bazakos et al., 2023). For wild olives, Unver et al. (2017) has provided the only to date available reference genome, while Zunino et al. (2023) employed a targeted sequencing approach to characterize 400 trees, mainly from the Western Mediterranean basin. They uncovered high genetic differentiation between some wild populations and cultivars, which implied the presence of genuine wild olives. Nevertheless, they highlighted several previously thought genuine wild populations as admixed, even within remote sites away from oleiculture activity. In another recent work, Mariotti et al. (2020) employed chloroplast and nuclear markers to analyse multiple genera of cultivated, ancient and wild genotypes among *O. europaea* and the Oleaceae family. Using population and comparative genomics they hinted of upcoming opportunities for uncovering the origin of agronomic and ecological traits of importance.

In order to prevent a possible genetic erosion of the olive species, research, evaluation and conservation of its genetic resources are of the utmost importance. A thorough analysis of the genetic wealth of olive genotypes will lead to new approaches to deciphering olive domestication and shed light on the distribution of genetic diversity (Baldoni et al., 2006). Furthermore, the estimation and comparison of the genetic diversity of wild and cultivated olives targeted at specific areas is the key to a better understanding of the reduction in genetic diversity resulting from the domestication and intensive use of agricultural practices.

The present study examines the level of genetic diversity in wild olive trees and its distribution in the north-east Mediterranean basin, an area of intense olive cultivation where wild and cultivated forms have long coexisted (Terral et al., 2004). In addition, cultivar samples from reference collections and from the field were uniquely genotyped to: 1) perform a comparative analysis of the genetic diversity of wild and cultivated olive trees; 2) shed light on genetic structure patterns within wild populations and 3) between wild and cultivated trees, and 4) investigate the occurrence of gene flow between these gene pools.

2 Materials and methods

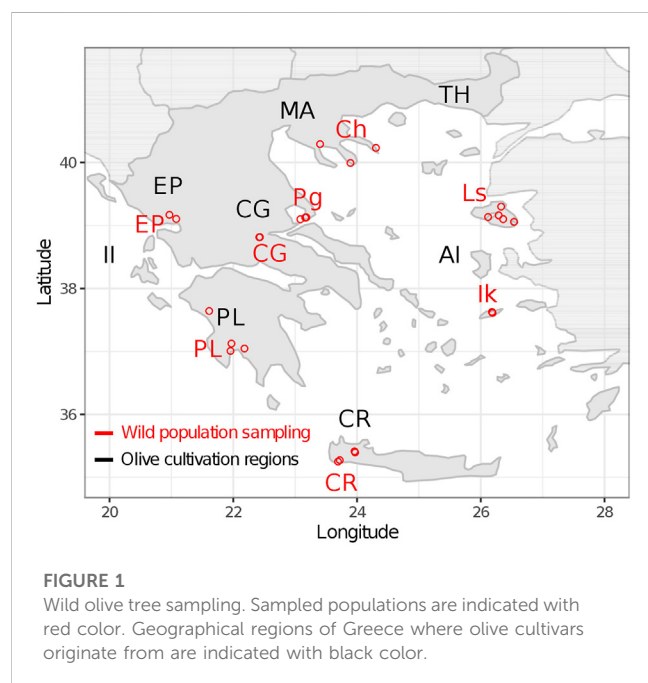
2.1 Sampling and DNA extraction

A total of 219 wild olive tree samples were collected from eight geographical regions in Greece (Table 1). In every region, multiple sampling plots were utilized as extended populations were unavailable (Figure 1). We also hypothesized that spreading our sampling effort would increase the chances of collecting samples with low genetic affinity to cultivars. The sample distance was at a minimum of 20 m. Young leaves were collected from individuals with the oleaster phenotype (trees or shrubs, often multi-stemmed with spinescent branches). Where wild populations were unavailable, root suckers from centennial cultivar trees were sampled. For the Chalkidiki population, two oleaster samples were also received from the herbarium of the School of Forestry and Natural Environment, Thessaloniki Aristotle University - TAUF. These were collected in the area of Agion Oros in 1999. Coordinates for the sampling plots of each population are provided in Supplementary Table S1.

TABLE 1 Wild and cultivated olive tree sample origin.

Gene pool	Sample Origin	Sample Origin abbreviation	Geographic region	Geographic region abbreviation	Number of samples
Wild olive trees (W)	Crete	CR	Crete	CR	30
	Chalkidiki	Ch	Macedonia	MA	33
	Lesvos	Ls	Aegean Islands	AI	30
	Peloponnese	PL	Peloponnese	PL	29
	Epirus	EP	Epirus	EP	30
	Central Greece	CG	Central Greece	CG	30
	Ikaria	Ik	Aegean Islands	AI	30
	Pagasetikos Bay	Pg	Aegean Islands	AI	7
Cultivars (C)	NOGB Reference samples	C	^a	^a	46
	Greek cultivars (Field collection)		^a	^a	14
	Monumental tree “Throuba Naxou” and commercial clone “Throuba”		Aegean Islands	AI	2
	Spanish cultivars		Spain	ES	2
Outgroup (OUT)	<i>Olea europaea</i> subsp. <i>cuspidata</i>	OUT	Asia/Africa	OUT	3
Total					286

^asee Table 2.



Forty six reference Greek cultivar leaf samples were provided by the National Olive Germplasm Bank of Greece (NOGB), Institute of Olive Tree, Subtropical Crops and Viticulture (IOSV), Hellenic Agricultural Organization “ELGO - Dimitra” (Chania, Southern Greece). This collection was enriched by 14 additional samples from olive groves nearby wild samples,

one commercially available clone “Throuba”, two Spanish cultivars (“Arbequina”, “Picual”), a monumental tree from Naxos island with an estimated age of over 3,000 years (Bazakos et al., 2023) and three *O. europaea* subsp. *Cuspidata* accessions as outgroups (Table 1, Table 2). Plant tissue was stored at -80°C until DNA extraction with the DNeasy Plant Pro Kit (Qiagen Inc., Valencia, CA).

2.2 Microsatellite genotyping

A set of 11 microsatellite markers (Sefc et al., 2000; Carriero et al., 2002; Cipriani et al., 2002; Díaz et al., 2006) was selected for genotyping using i) reproducibility and ii) ease of scoring as criteria (Doveri et al., 2008; Baldoni et al., 2009). A touchdown PCR procedure was used: 95°C for 5 min, 10 touchdown cycles: 95°C for 20 s, 60°C for combination A and 65°C for combination B for 30 s (-1°C step/cycle) and 72°C for 30 s. It was followed by 25 cycles: 95°C for 20 s, 50°C for combination A and 54°C for combination B for 30 s, 72°C for 30 s, final extension step at 72°C for 10 min. Fragment analysis was conducted in an ABI 3730xl (Applied Biosystems, Foster City, CA, United States) with GeneScan 500 LIZ size standard and results were recorded with GeneMapper v4. Cultivar reference samples were genotyped twice to avoid sample mix-up (two independent DNA extractions and PCR amplifications). PCR amplifications were repeated for a random sample of 30 individuals ($\sim 10\%$ of the whole data set). These data were used to calculate i) error rate per reaction and ii) per allele (Hoffman and Amos, 2005).

TABLE 2 Principal area of cultivation of olive cultivars investigated in the present study.

Cultivar	Principal area of cultivation	Region	Region abbreviation
ADRAMYTINI	Lesvos island, Andros island	Aegean Islands	AI
AGOUROMANAKOLIA	Kynouria, Argolida, Korinthia	Peloponnese	PL
AMYGDALOLIA	Attika, Evia, Voeotia	Central Greece	CG
ASPROLIA ALEXANDROUPOLIS	Alexandroupoli	Thrace	TH
ASPROLIA LEFKADOS	Lefkada and Meganissi islands, Aetoloakarnania	Ionian Islands - Western Central Greece	II - CG
CHALKIDIKIS	Chalkidiki	Macedonia	MA
CHONDROLIA CHALKIDIKIS	Chalkidiki	Macedonia	MA
CHRYSLIA	Outgroup	Outgroup	OUT
DAFNELIA	Islands of Chios, Samos, Ikaria, Andros, Siros, Naxos, Sifnos (all Central and Eastern Aegean islands)	Aegean Islands	AI
FRANTOIO RODOU	Rhodos island and Italy (Italian origin)	Aegean Islands	AI
GAIDOURELIA	Kynouria	Peloponnese	PL
GALATISTAS AG.OROUS	Agion Oros and Monopigado	Macedonia	MA
KALAMON	Lakonia, Messinia, Ilia, Aetoloakarnania, Fthiotida Kynoyria	Peloponnese	PL
KALOKERIDA	Kerkira island	Ionian Islands	II
KAROLIA	Lesvos island	Aegean Islands	AI
KARYDOLIA	Voeotia, Fokida	Macedonia	CG
KOLYBADA	Attika and Tinos island	Aegean Islands	AI-CG
KOLYREIKI	Ilia	Peloponnese	PL
KONSERVOLIA/AMFISSIS	Igoumenitsa, Magnissia, Evia, Fthiotida, Aetoloakarnania, Fokida, Achaia	(Western) Central Greece - Ionian Islands	CG-II
KORONEIKI	Crete, West and South Peloponnese, Central and South Aegean and Ionian islands	Crete - Peloponnese	CR-PL
KOTHREIKI	Argolida, Fokida, Voeotia, Korinthia, Fthiotida	Central Greece	CG-PL
KOUTSOURELIA	Korinthia, Lakonia and Aetoloakarnania	Central Greece - Peloponnese	CG-PL
LEFKOLIA SERRON	Serres	Macedonia	MA
LIANOLIA KERKYRAS	Kerkyra and Paxi island, Igoumenitsa and Preveza	Ionian Islands	II
LIANOMANAKO TYROY	Kynouria	Peloponnese	PL
MAKRIS	Evros, Rodopi and Samothraki island	Thrace	TH
MASTOIDIS/TSOUNATI	Crete, Lakonia, Messinia, Arkadia, Ilia and Achaia	Crete - Peloponnese	CR-PL
MATOLIA ILIAS	Ilia	Peloponnese	PL
MAVRELIA	Messinia	Peloponnese	PL
MAVRELIA SERRON	Serres	Macedonia	MA
MAVROLIA LEFKADOS	Lefkada island	Ionian Islands	II
MEGAREITIKI	Attika, Argolida, Voeotia, Fthiotida, Fokida, Kynouria, Achaia	Central Greece - Peloponnese	CG-PL
MYRTOLIA	Lakonia	Peloponnese	PL
NTOPIA ZAKYNTHOU	Zakynthos island	Ionian Islands	II
PETROLIA	Serres	Macedonia	MA

(Continued on following page)

TABLE 2 (Continued) Principal area of cultivation of olive cultivars investigated in the present study.

Cultivar	Principal area of cultivation	Region	Region abbreviation
PIERIAS	Pieria	Macedonia	MA
PIERIAS (SKOTINIS)	Pieria	Macedonia	MA
PIKROLIA	Kerkira island	Ionian Islands	II
RACHATI	Achaia and Aetoloakarnania	Peloponnese - Central Greece	CG-PL
STROGGYLOLIA	Chalkidiki	Macedonia	MA
THIAKI	Kefalonia and Ithaki islands	Ionian Islands	II
THROUBA THASOU	Thassos island, Kavala	Aegean Islands	AI-MA
THROUBOLIA	Crete (Rethymno, Herakleion), Evia and many Aegean islands	Aegean Islands - Crete	AI-CR
TRAGOLIA	Messinia, Lakonia	Peloponnese	PL
VALANOLIA	Lesvos, Chios and Skyros islands	Aegean Islands	AI
VASILIKADA	Kerkira island, North Evia	Central Greece	CG
Ls-C-N-01	Lesvos	Aegean Islands	AI
Ls-C-N-02	Lesvos	Aegean Islands	AI
Ls-C-N-03	Lesvos	Aegean Islands	AI
Ls-C-N-04	Lesvos	Aegean Islands	AI
Ls-C-N-05	Lesvos	Aegean Islands	AI
Ls-C-N-06	Lesvos	Aegean Islands	AI
Ls-C-N-07	Lesvos	Aegean Islands	AI
Ik-C-N-01	Ikaria	Aegean Islands	AI
Ik-C-N-02	Ikaria	Aegean Islands	AI
Ik-C-N-03	Ikaria	Aegean Islands	AI
Ik-C-N-04	Ikaria	Aegean Islands	AI
CR-C-N-01	Crete	Crete	CR
CR-C-N-02	Crete	Crete	CR
CR-C-N-03	Crete	Crete	CR
THROUBA	Crete (Rethymno, Herakleion), Evia and many Aegean islands	Aegean Islands	AI
THROUBA_NAXOU	Naxos island	Aegean Islands	AI
ARBEQUINA	Spain	Spain	ES
PICUAL	Spain	Spain	ES
OUT-1	-	Outgroup	OUT
OUT-3	-	Outgroup	OUT
OUT-4	-	Outgroup	OUT

2.3 Data analysis

Statistical analysis was performed in R 4.1.3 (R Core Team, 2021). Basic population statistics were estimated with poppr (Kamvar et al., 2014), hierfstat (Goudet, 2005), adegenet (Jombart, 2008), and genepop (Rousset, 2008). Synonymy-clonality (different samples with identical multilocus

genotypes-MLG) was determined in poppr. To avoid cases of samples differing only due to genotyping error and mutations in hyper-variable loci, MLGs were collapsed by dissimilarity distance (denoted in the rest of the text as “contracted MLGs”). The minimum dissimilarity distance at which two individuals would be considered as “contracted MLGs”, was decided by identifying an initial small peak in the histogram

of pairwise distances, as suggested by Arnaud-Haond et al. (2007). Due to the limited sample size, the population of Pagasitikos Bay was not evaluated for population level inferences. To preserve custom code produced for this analysis a new R package was developed named *PopGenUtils* (<https://github.com/nikostourvas/PopGenUtils>). It was used to calculate the probability of identity (PID) (Paetkau and Strobeck, 1994) and the more conservative statistic (PIDsibs) (Taberlet and Luikart, 1999). Polymorphic information content was estimated with polysat (Clark and Jasieniuk, 2011).

$$PID = 2\left(\sum p_i^2\right)^2 - \sum p_i^4$$

$$PIDsibs = 0.25 + \left(0.5\sum p_i^2\right) + \left[0.5\left(\sum p_i^2\right)^2\right] - \left(0.25\sum p_i^4\right)$$

where p_i is the frequency of allele i .

Rarefaction curves were produced with *vegan* (Oksanen et al., 2020). Genetic structure was evaluated using G_{ST} as calculated in *strataG* (Archer et al., 2017). The admixture and the correlated allele frequency models (Falush et al., 2003) of *STRUCTURE* 2.3.4 software (Pritchard et al., 2000) via the multi-core implementation provided by *Structure_threader* (Pina-Martins et al., 2017), were used. Wang's (2017) suggestions to avoid bias due to asymmetrical sampling were implemented. In particular, i) the alternative ancestry prior was used by setting $POPALPHAS = 1$ and ii) the default $ALPHA$ value 1) was reduced to $ALPHA = 0.2$ based on the suggestion of $1/K$ (where K is the expected number of groups in the data set). Burn-in period was set to 5×10^5 , followed by 10^6 iterations for $K = 1$ up to $K = 8$ with 20 repeats for each K value. The "distruct for many K's" pipeline from *CLUMPAK* (Kopelman et al., 2015) was implemented to aggregate results. Outgroup samples were not used in genetic structure analysis as they represent an extreme case of asymmetric sampling (Neophytou, 2014; Puechmaille, 2016). Membership proportions were visualized with *pophelper* (Francis, 2017).

AssignPOP 1.2.4 (Chen et al., 2018) was used to perform assignment tests with the SVM prediction algorithm. This software splits the given data set in "training" and "test" data based on cross-validation (in this case 100 iterations). A dimension reduction step follows, using PCA analysis and a machine learning algorithm classifies the samples. The developed predictive model is evaluated using the withheld "test" data which helps avoid high-grading bias (Anderson, 2010).

The presence of one up to five gene flow events between gene pools was investigated with *TreeMix* (Pickrell and Pritchard, 2012), with gene pools being determined based on *Structure* results. Ten independent runs were executed for each evolutionary scenario. Data input for *TreeMix* was prepared by calculating the mean and the standard deviation of the absolute length of each microsatellite. The function *genind2treemix* is provided in the *PopGenUtils* package to produce such an input file. Code from the R package *OptM* (Fitak, 2021) was used to plot mean likelihood values for different scenarios of gene flow events.

The complexity of the analyses offered by modern statistical tools has led to problems of reproducibility of research results (Gilbert et al., 2012). To counter that, all statistical analyses were conducted in a container-based computing environment (https://github.com/nikostourvas/forest_popgen), which included the

operating system, software, scripts and raw data following the recommendations of Marwick et al. (2018). This was registered in a Zenodo digital repository (10.5281/zenodo.8104610) and is available to any party interested in regenerating and verifying the results of this work.

3 Results

3.1 Data quality check and clone detection

Locus GAPU103A displayed suboptimal PCR amplification and was dropped from analyses. Following the PCR repeatability test in 30 individuals, 10 errors were recorded (error rate per reaction: 1.69%). All cases were with regards to erroneous scoring of an allele (error rate per allele: 0.85%) (Table 3). Polymorphic Information Content (PIC) varied from 0.926 for marker DCA09 to 0.735 for marker DCA13. To assess whether the remaining 10 microsatellite loci provided sufficient statistical power to distinguish olive tree samples and elucidate patterns of fine genetic structure, a statistical resampling procedure was executed. PID and PIDsibs values were calculated by sampling with replacement over loci 1,000 times for 1–9 loci (Supplementary Table S2, Supplementary Figure S2A). The same procedure was followed for calculating the number of detected MLGs for different combinations of the genotyped loci (Supplementary Figure S2B). Random combinations of as few as three loci were enough to achieve low PID (median PID = 2.12×10^{-5}) and discriminate most MLGs. The overall probability of identity (PID) was 1.60×10^{-16} , while overall PIDsibs was 1.25×10^{-5} .

Five cases of synonymy-clonality were detected (Table 4), all of which concerned cultivar samples and two of those concerned the NOGB. The latter were the sample pairs "Koutsourelia" - "Rachati", and "Pierias (Skotinis)" - "Pierias". The rest concerned cultivar samples from Ikaria which were identified as "Koroneiki" cultivar and one cultivar sample from Lesvos which was identified as "Chondrolia Chalkidikis". Two cultivar samples from Crete had identical multilocus genotype. Additionally, a histogram of genetic distances between samples (each unit of distance equals one allele of difference) displayed a small peak at 1–3 distance units (Figure 2). The individuals forming this peak had 17–19 common alleles out of 20; hence they were regarded as "contracted MLGs" following the "contracted" multilocus genotype analysis available in *poppr*. These six contracted MLGs mostly involved cultivars as well. The reference sample "Chrysolia" had a match with one outgroup and reference samples "Chondrolia Chalkidikis" and "Chalkidikis" matched at 19/20 alleles. Two samples from olive groves in Lesvos were identified as "Valanolia" and one more sample from Lesvos was matched to "Adramytini" reference. Based on the overall results, the three samples from olive groves in Crete were all matched to "Mastoidis/Tsounati". Finally one putative wild sample from Central Greece, two cultivar samples from Ikaria, reference sample "Throubolia", commercially available "Throuba" and the monumental tree "Throuba Naxou" formed the last contracted MLG. The putative wild sample was removed from all further analyses. Only one sample from each MLG was included in analyses of genetic structure to avoid bias. There remained three cases of samples from olive groves of Lesvos (Ls-C-N-01, Ls-C-N-04,

TABLE 3 Error rate per reaction/allele and Polymorphic information content for the 10 microsatellite loci analyzed.

Locus	Number of reactions	Number of errors per reaction	Number of errors per allele	Error rate per reaction	Error rate per allele	PIC
DCA09	59	3	3	0.05	0.025	0.926
UDO43	59	2	2	0.0333	0.01667	0.913
DCA16	59	1	1	0.0167	0.0083	0.895
DCA18	60	2	2	0.0333	0.0167	0.889
DCA14	60	0	0	0	0	0.878
IAS-oli23	60	0	0	0	0	0.868
DCA03	60	1	1	0.0167	0.0083	0.837
GAPU71B	54	1	1	0.0167	0.0083	0.818
DCA05	60	0	0	0	0	0.785
DCA13	59	0	0	0	0	0.735
Mean				0.0169	0.0085	0.854
Total	590	10	10			

TABLE 4 Synonyms and contracted MLGs among wild and cultivated olive samples.

Synonyms						
1	Ik-C-N-02	Ik-C-N-04	Koroneiki			
2	Ls-C-N-02	CHONDROLIA CHALK				
3	PIERIAS (SKOTINIS)	PIERIAS				
4	KOUTSOURELIA	RACHATI				
5	CR-C-N-01	CR-C-N-02				
Contracted MLGs						
1	CHRYSLIA	OUT-1				
2	CHONDROLIA CHALK	CHALKIDIKIS				
3	Ls-C-N-03	Ls-C-N-06	VALANOLIA			
4	Ls-C-N-05	ADRAMYTINI				
5	CR-C-N-02	CR-C-N-03	MASTOIDIS_TSOUNATI			
6	CG-W-R-02	Ik-C-N-01	Ik-C-N-03	THROUBOLIA	THROUBA	THROUBA NAXOU

Ls-C-N-07) for which it was not possible to match them to a cultivar of the national collection.

3.2 Wild population and cultivar genetic diversity, inbreeding and genetic differentiation

For wild populations in the 10 loci studied, 211 alleles were observed. The number of alleles per locus ranged from 15 (DCA13) to 30 (DCA16), with a mean of 21.1. The effective number of alleles ranged from $ne = 4.558$ (DCA13) to $ne = 14.535$ (DCA09), with a mean of $ne = 9.109$. Observed heterozygosity ranged from 0.396

(DCA13) to 0.885 (IAS.oli23), and gene diversity (Nei, 1978) ranged from 0.780 (DCA13) to 0.931 (DCA09) (Supplementary Table S3).

All wild populations had mean allelic richness (AR) values > 10 , with the lowest value observed in the population of the Peloponnese ($AR = 10.65$) and the highest in the population of Lesvos ($AR = 14.05$). Conversely, the population of Lesvos exhibited the lowest observed heterozygosity value ($H_O = 0.695$), while the highest was recorded in the population of Crete ($H_O = 0.825$). All loci presented positive values of the inbreeding coefficient F_{IS} , except for the one corresponding to marker IAS. oli23 (-0.001). The mean F_{IS} value was 0.156. The 95% confidence intervals for the inbreeding coefficient F_{IS} were estimated to range above zero for the populations of Central Greece, Chalkidiki, Ikaria, and Lesvos.

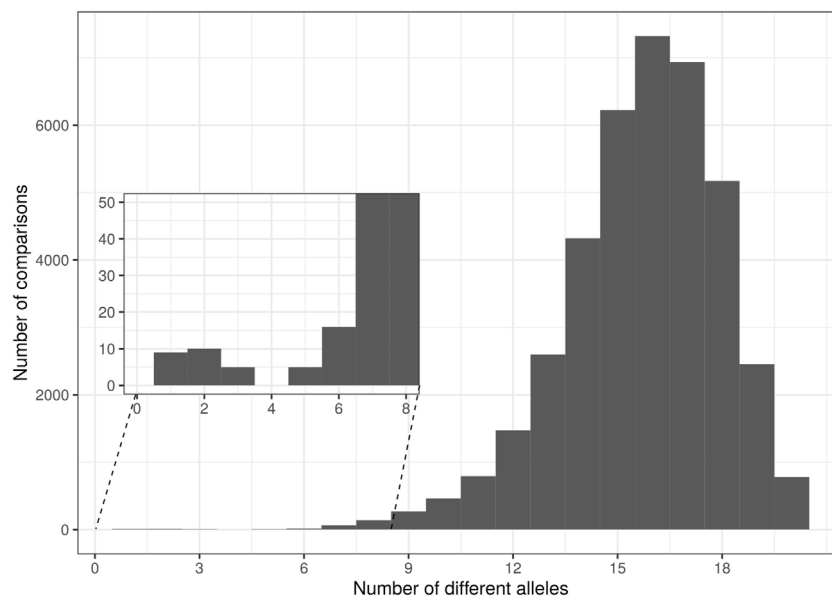


FIGURE 2
Histogram of pairwise genetic distances among all investigated olive tree samples.

Twenty-three private alleles were detected among wild populations. Only the population of the Pagasitikos Bay did not carry any private allele, while eight were identified in the population of Ikaria (Figure 3).

Genetic distances among the wild populations ranged at low levels, with the lowest G_{ST} value detected between the populations of Lesvos and Central Greece ($G_{ST} = 0.001$, $p = 0.220$) and the highest between Lesvos and Peloponnese ($G_{ST} = 0.032$, $p < 0.001$). Due to the expected behaviour of G_{ST} to present low values when calculated from hyper-variable markers such as microsatellites, its standardized version G'^{*}_{ST} (Meirmans and Hedrick, 2011) and Jost's D (Jost, 2008) are also available in Supplementary Figure S1.

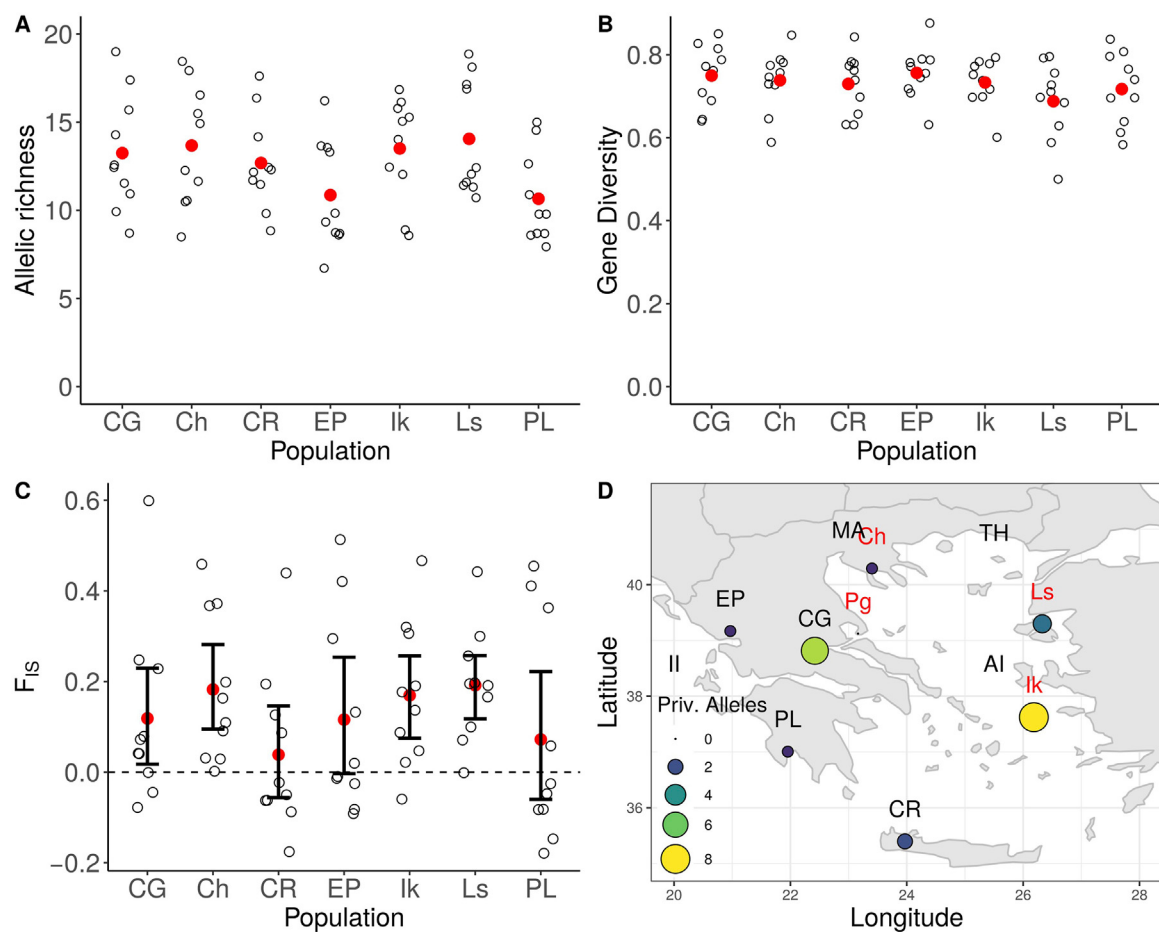
Cultivated varieties displayed 127 alleles and the number of alleles per locus ranged from seven (DCA13, DCA14) up to 20 (UDO43). The lowest value for effective number of alleles $n_e = 2.175$ was observed for marker DCA05 and the highest $n_e = 10.520$ for DCA09, while the mean value was $n_e = 6.313$. Mean observed heterozygosity $H_o = 0.787$ was slightly higher than the overall value detected in wild populations. Conversely, gene diversity $uHe = 0.806$ was lower than the corresponding value in wild olive trees (Supplementary Table S4).

3.3 Genetic differentiation between wild and cultivated olive trees

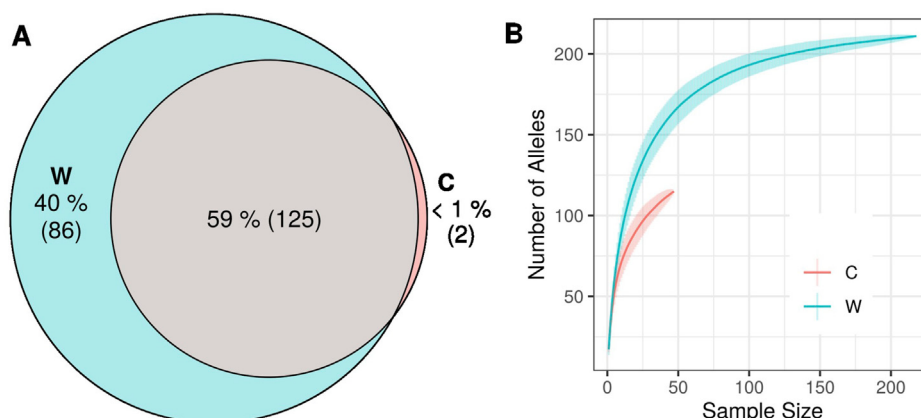
Two alleles (<1% of all alleles) were exclusively found in cultivars, while 86 alleles (40% of the total alleles) were only present in wild populations (Figure 4A). Rarefaction curves showed a substantially higher number of alleles for wilds compared to cultivars, given similar sample sizes (Figure 4B). The genetic differentiation between wild genotypes and Greek cultivars, was low ($G_{ST} = 0.014$) but statistically significant ($p <$

0.001). The complementary differentiation measures $G'^{*}_{ST} = 0.175$, $p < 0.001$ and $D = 0.151$, $p < 0.001$ were also recorded.

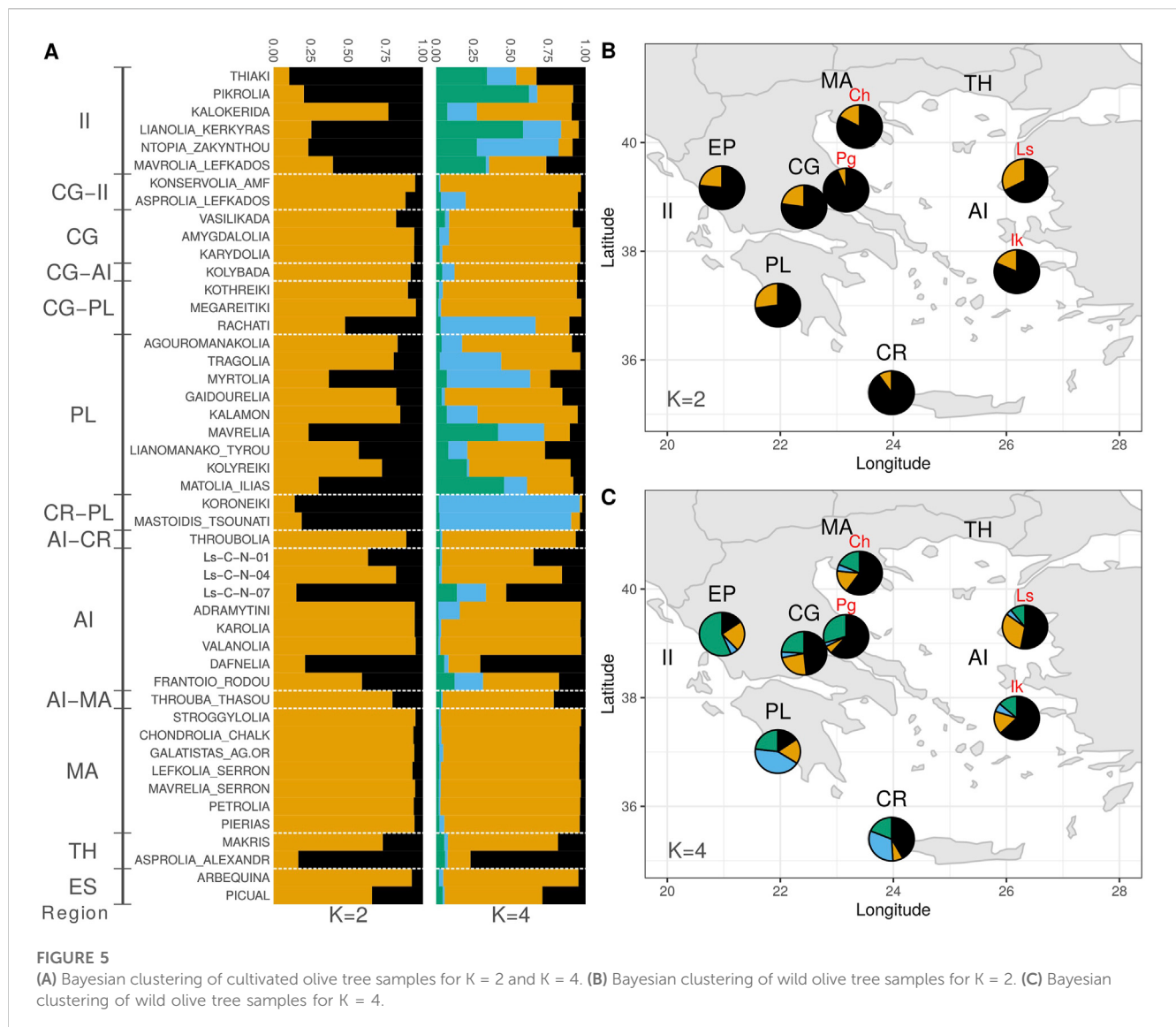
For a better understanding of the genetic structure, Bayesian clustering was employed via the STRUCTURE software. Because locus DCA13 exhibited Hardy-Weinberg disequilibrium in six out of the seven populations tested, the analysis was performed with and without the presence of this locus. Due to the similarity of the results of both analyses, data obtained from all available loci is presented. The optimal value of K groups was decided based on plotting the average of the estimated natural logarithm of the probability of the data indicating values of $K = 3$ and $K = 4$. The Evanno method (Evanno et al., 2005) indicated a value of $K = 2$ (Supplementary Figure S3). For $K = 2$, the analysis indicated one group in which the majority of the cultivars were classified and another group containing the largest subset of wild genotypes (Figure 5). For $K = 4$, results demonstrated geographic coherence. A new group appeared, which was consistent with the geographical area of Southwestern Greece (light-blue color), with the main representatives being the cultivars "Koroneiki" and "Mastoidis/Tsounati", and high membership coefficients in the cultivars "Ntopia Zakynthou", "Myrtolia", "Rachati" and "Tragolia". In wild genotypes, the largest subset of samples in the Peloponnese and Crete participated in the same group. Furthermore, another group was formed, which was mainly associated with the geographic area of Western Greece (green color). Specifically, high membership percentage was observed in two cultivars of the Ionian Islands ("Pikrolia", "Lianolia Kerkyras") and lower percentages in the cultivars "Thiaki", "Mavrelia" and "Matolia Ilias". Almost all samples of the wild population of Epirus and some samples from other geographic regions showed high similarity to this group. Most of the cultivars from the regions of Macedonia, Central Greece, the Aegean Islands and Thrace, as well as some cultivars of the Ionian Islands and Peloponnese, made up the most populous group of

**FIGURE 3**

Genetic diversity statistics for wild olive tree populations. (A) Allelic richness, (B) Gene Diversity, (C) Inbreeding coefficient F_{IS} , (D) Geographical distribution of private alleles.

**FIGURE 4**

(A) Venn diagram of total number of alleles detected in the data set partitioned between wild and cultivated olive tree samples. (B) Rarefaction curves of number of alleles in wild and cultivated samples.



cultivars (orange color). It is worth noting that Spanish cultivars "Arbequina" and "Picual" had high membership proportions for the same group, similarly to some wild genotypes from Lesbos, Central Greece, Chalkidiki and Ikaria. Finally, the cultivars "Asprolia Alexandroupolis" and "Dafnelia" showed high participation rates in the last group of genotypes (black color), which included a large number of wild samples from the regions of Lesbos, Ikaria, Chalkidiki, Central Greece, and Pagasitikos Bay, forming a cluster that corresponds to the geographic area of the Aegean Sea.

The subsequent assignment analysis was conducted with the following predefined groups based on STRUCTURE results: i) wild samples or cultivars, ii) Peloponnese and Crete wilds or North-Eastern Greece and Epirus wilds, and iii) Peloponnese and Crete wilds or Epirus wilds or North-Eastern Greece wilds. To avoid bias related to unequal sample sizes, care was taken to use a balanced "training" data set (e.g., for the first comparison, 45 cultivars and 48 wild samples - six from each population - were used). On average 83% of wild genotypes and 68% of cultivars were assigned to the correct gene pool ("wild" or "cultivars") (Figure 6A). When only wild individuals were taken into account and were separated into

two groups "Peloponnese and Crete" (PL and CR) and "Northern-Eastern Greece and Epirus" (NE-GR) according to the STRUCTURE analysis results for $K = 3$, on average 69% of genotypes with Peloponnese or Crete origin and 69% of the remaining wild individuals were assigned to the correct group. These percentages increased to 73% and 71% respectively, when three random loci were used (Figure 6B). In the case of separating wild samples into three groups, 61% of individuals were correctly assigned to the population of Epirus (EP), 64% were correctly assigned to the group formed by the populations of Peloponnese and Crete (PL and CR), and 57% to the group of Northern-Eastern Greece (NE-GR) (Figure 6C).

3.4 Gene flow

To investigate the demographic history of the identified gene pools, a maximum likelihood tree was constructed using the TreeMix software. Similarly to the assignment analysis, populations were grouped based on the Bayesian clustering

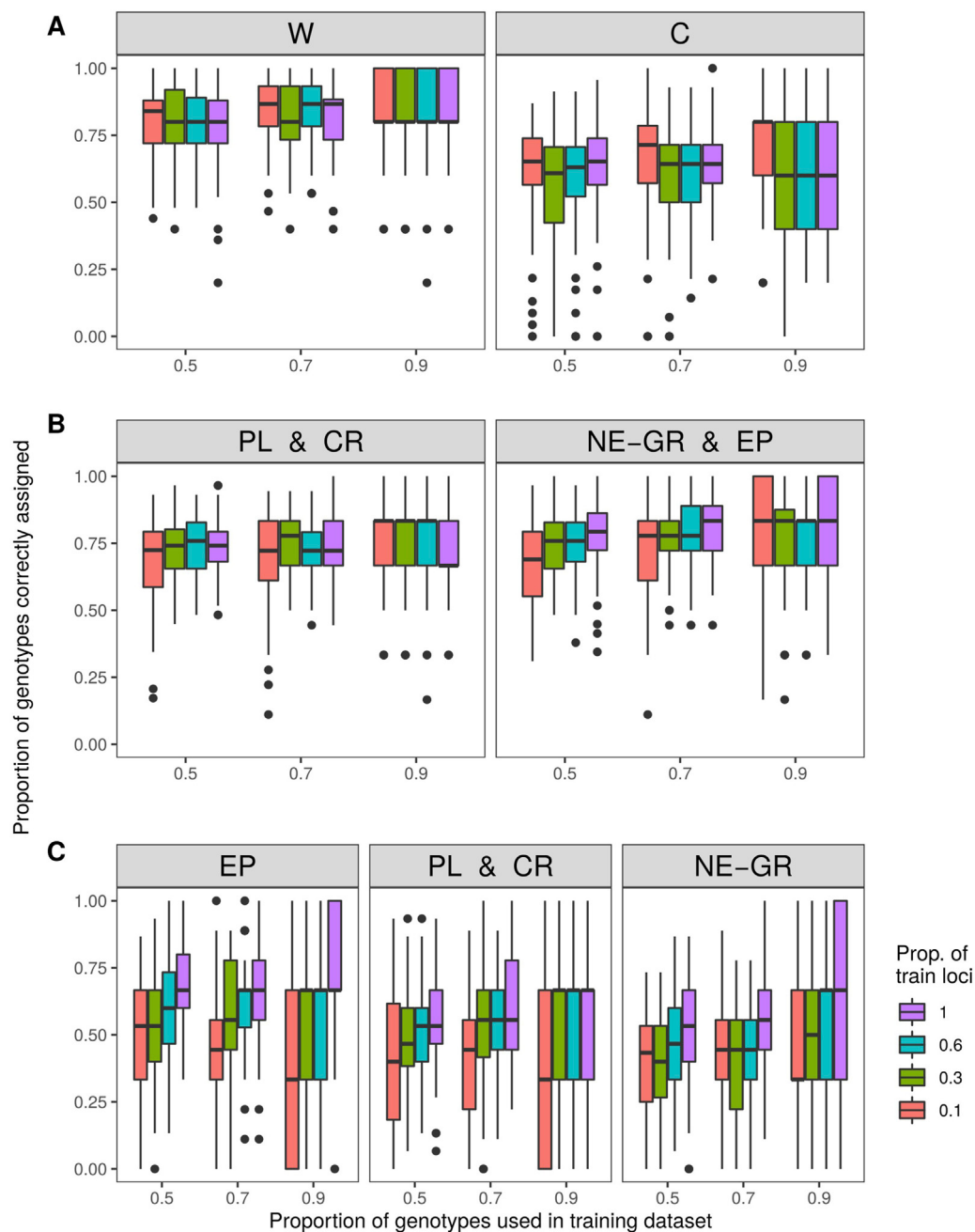


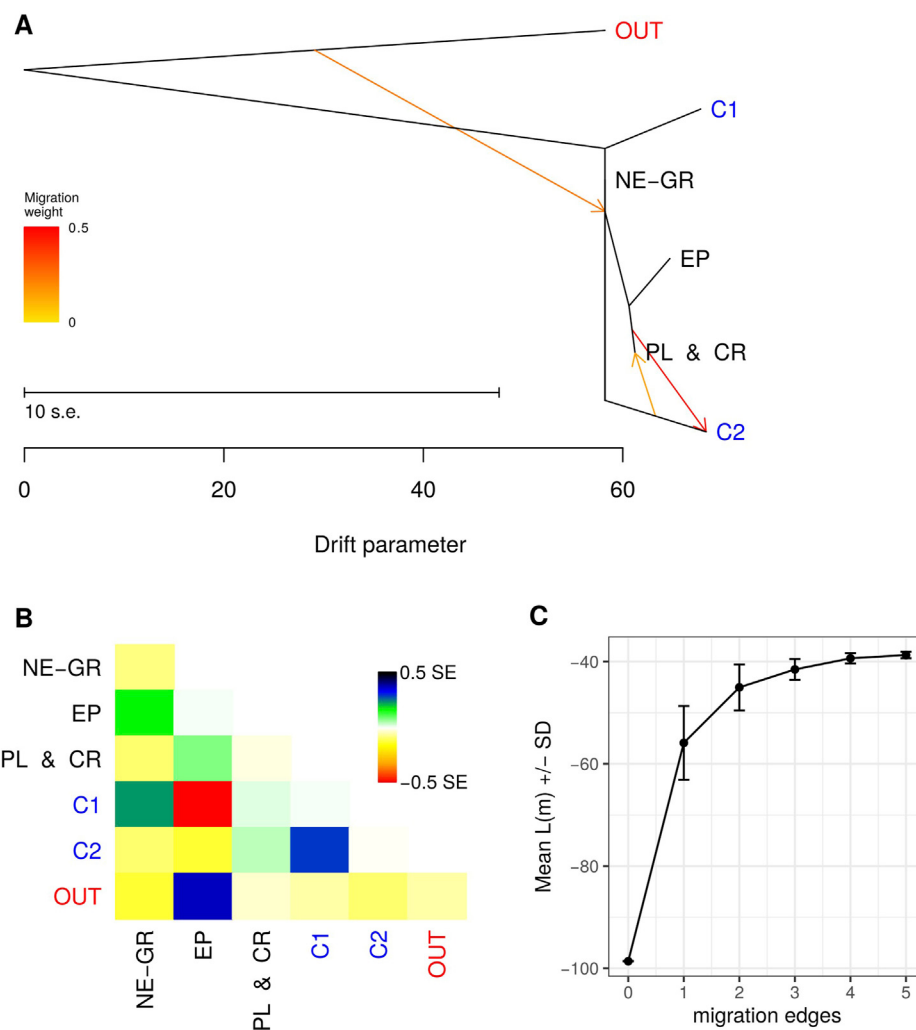
FIGURE 6

Assignment accuracy after 100 iterations with three levels of training samples (50%, 70%, 90% of samples) and four levels of training loci (1, 3, 6, 10 loci). Investigated groups: (A) wild vs. cultivated olive tree samples, (B) within wild olive tree samples from Peloponnese and Crete vs. North-Eastern Greece and Epirus, and (C) within wild olive tree samples from Epirus vs. Peloponnese and Crete vs. North-Eastern Greece.

output. The wild olive genotypes were classified into three groups: "Northeastern Greece" (NE-GR), "Peloponnese-Crete" (PL and CR), and "Epirus" (EP). Most cultivars were included in a separate group (C1 - orange color in the STRUCTURE analysis), while the remaining cultivars, which had a membership percentage of <50% in group C1, were included in a second group (C2) since they mostly constituted "mosaics" of the three groups wild samples were members of (Supplementary Table S5).

Based on the progression of likelihood values for different numbers of migration events (Figure 7C), the three-event model

was selected. The 10 independent repetitions of the model did not yield identical results, nevertheless the dendrogram presented in Figure 7A resulted from 5 repetitions and attributed the relationships among the gene pools with the lowest standard errors compared to the remaining iterations. The first migration event indicates that 24% of the origin of the populations in Northeastern Greece (NE-GR) originates from a parental population that resembled genetically *O. europaea* subsp. *Cuspidata*. According to the second detected gene flow event, the C2 cultivar group emerged from gene transfer

**FIGURE 7**

(A) Maximum Likelihood dendrogram of wild and cultivated olive tree sample groups as indicated by Bayesian clustering with three gene flow events inferred by TreeMix. (B) Residuals of the TreeMix model. (C) Mean likelihood values for none up to five gene flow events.

(migration intensity $w = 48\%$) from populations in Peloponnese and Crete (PL and CR). Finally, a third incident of gene flow with the opposite direction compared to the second event, but with distinctly lower intensity ($w = 15\%$).

4 Discussion

4.1 Quality check and statistical power

Error rates per reaction (average 1.69%) and per locus (average 0.85%), were at sufficiently low levels, ensuring the reproducibility of molecular analyses. In a similar study aimed at compiling a list of recommended loci for genotyping olive cultivars, Baldoni et al. (2009) selected 11 SSRs from a set of 37, after conducting molecular analysis that was independently repeated in four laboratories. For the 11 suggested loci (eight of which were also used in this study), the error rate per locus was calculated to be 2.26%. As for the frequency of errors among loci, it was observed

that the most informative markers according to the PIC index, also had higher error rates. This phenomenon is previously reported (Flores-Renteria and Krohn, 2013) and has been observed in other species such as grapevine (Cipriani et al., 2008). One proposal for minimizing genotyping errors is to avoid highly polymorphic dinucleotide markers as they often come with issues related to false peaks, in favour of microsatellites with repeat motifs of three to five nucleotides (Dewoddy et al., 2006).

The PID (1.60×10^{-16}) and PIDsibs (1.25×10^{-5}) indices showed low values, confirming the expected high discriminatory power to identify cases of clonality within the dataset. Compared to other studies, Haouane et al. (2011) genotyped the worldwide olive genetic bank of Marrakech, Morocco (OWGB) and recorded an overall PID of 2.55×10^{-14} . This value was calculated using 12 microsatellite loci and 210 associated alleles. In another application involving 431 samples from wild and cultivated trees in the Spanish region of Andalusia, with 14 SSR loci and 191 alleles, a value of $\text{PID} = 1.96 \times 10^{-16}$ was estimated (Díez et al., 2011), which is in the same order of magnitude as this study. Although Haouane et al. (2011) detected

more alleles than [Diez et al. \(2015\)](#), the identity index was higher in the latter case. This is probably due to the fact that allelic frequencies between wild and cultivated gene pools differ enough, so that fewer alleles are required to distinguish samples with the same level of confidence. In addition, the discriminatory power of genetic markers is likely to be lower for cultivar samples which bear to some extent the imprint of genetic bottlenecks as a result of the domestication process ([Diez et al., 2015](#)).

4.2 Clone detection

Eleven out of 14 samples from local olive groves in Crete, Lesvos, and Ikaria, which were adjacent to the sampling areas of wild genotypes, were successfully matched to known cultivars. Regarding the reference samples, the cultivar pairs "Pierias" - "Pierias (Skotinis)", "Koutsourelia" - "Rachati", and "Chondrolia Chalkidikis" - "Chalkidikis", displayed identical, or highly similar multilocus genotypes, and were treated as clones. However, they possess morphological differences ([Koubouris et al., 2019](#)) and it is hypothesized that one member of each pair originated from undetected mutations from the other. Similar cases have also been suggested for cultivar pairs maintained at the World Olive Germplasm Bank of Cordoba ([Trujillo et al., 2014](#)). It is worth mentioning, that in the recent work by [Bazakos et al. \(2023\)](#) employing whole genome re-sequencing, the "Chondrolia Chalkidikis" - "Chalkidikis" pair had a proportion of genomes shared identical-by-descent $\text{PropIBD} = 0.9819$, while a $\text{PropIBD} = 0.8801$ was recorded for "Koutsourelia" - "Rachati".

Additionally, the "Chrysolia" accession was found to share the same contracted MLG with one of the outgroup samples and therefore should be classified as *O. europaea* subsp. *Cuspidata*. The reference sample of the national collection "Throubolia," the commercially available sample "Throuba" and the over 3,000 years old monumental olive tree "Throuba Naxou" formed another contracted MLG (also confirmed by [Bazakos et al., 2023](#)). "Throubolia" cultivar possesses the unique agronomic characteristic of fruit ripening and being instantly edible after harvesting ([Kostelenos and Kiritsakis, 2017](#)). Such a trait is expected to have been highly desirable during the domestication process and the identification of this monumental tree supports the hypothesis that the Greek geographic region served as an ancient hotspot for olive cultivar development. This hypothesis could hold true even if the Greek geographic region at large, did not act as a primary centre of olive domestication.

4.3 Genetic diversity within wild gene pool

All populations of wild olive trees showed high levels of genetic diversity, as evidenced by the high values of allelic richness and gene diversity, a result which is expected for a species that is characterized by a complex self-incompatibility system ([Koubouris et al., 2014](#); [Breton et al., 2021](#)). In all markers but IAS. oli23, a heterozygote deficiency was observed, which might be linked to small population sizes, limited gene flow between fragmented local populations and/or genetic bottlenecks.

The populations of Epirus and of the Peloponnese displayed slightly lower allelic richness values than the rest. In the populations of Ikaria and Central Greece, eight and seven private alleles were found respectively, while in the rest of the populations, one to three private alleles were recorded (excluding Pagasitikos Bay). This pattern may reflect the intensity of genetic pollution, with areas of historically lower oleiculture intensity maintaining a higher percentage of rare alleles. Undoubtedly, the search for wild individuals, whether true-to-type wild olives or ferals, should focus on geographical areas far from places with a long tradition of olive cultivation.

4.4 Genetic differentiation between wild and cultivated olive trees

The level of genetic differentiation between wild and cultivated olive gene pools, as expressed by the G_{ST} coefficient, was low although statistically significant ($G_{ST} = 0.014$, $p < 0.001$). Low levels of genetic differentiation between the two pools have also been proposed in most regions of the Mediterranean ([Belaj et al., 2007](#); [Hannachi et al., 2008](#); [Diez et al., 2015](#)). The absence of strong differentiation signals indicates that the collected wild olive samples should probably not be considered "true-to-type" wilds, but rather feral forms (i.e., individuals that have arisen from multiple admixture events between wild ancestors and cultivated genotypes). However, it should be noted that levels of genetic differentiation observed in the olive tree are lower compared to other perennial tree species, partly due to the long time between successive generations ([Besnard et al., 2018](#)). A proper characterization as wild or feral, would require reference samples of wild genotypes and employment of more phylogenetically informative marker systems, such as organelle microsatellites or high throughput sequencing.

Almost all (>99%) of cultivar alleles were detected in wild individuals. In contrast, 40% of the total alleles were recorded only in the wild populations. This difference in the number of private alleles was not a statistical artifact due to differential sampling effort, as confirmed by rarefaction curves which demonstrate that for similar sample sizes the cultivated gene pool withholds fewer alleles. Rarefaction curves provide valuable conclusions when samples are representative of each group. The stratified sampling over eight broad geographical areas for wild genotypes, ensures that this condition is met. Ensuring the same for cultivars, i.e., that accessions which make up the national collection constitute an accurate representation of the total genetic diversity that is found in the country's cultivated plant genetic material, is more challenging. However, it can be considered as a reasonable assumption since the origin of cultivars that are included in NOGB spatially covers all olive cultivation regions in Greece. This is the first study that reports the extent of the wealth of unique genetic variation present in wild olive gene pools in Greece. The finding indicates that wild samples cannot have emerged exclusively from cultivars escaping cultivation. It also corresponds to a genetic bottleneck as a result of domestication as previously recorded by [Besnard et al. \(2013b\)](#) who reported a loss of genetic diversity of around 16% (for nuclear loci) and 64% (for chloroplast loci) in cultivated olive genotypes compared to wild individuals. Similar

results have also been reported by Lumaret et al. (2004) where significantly lower heterozygosity was found in cultivars compared to wilds from the eastern and western Mediterranean.

The Bayesian analysis performed using the STRUCTURE software clustered together several cultivars with the same origin/principal area of cultivation, especially in the Peloponnese, Crete, and Epirus regions. Cultivars were not strictly clustered according to region though. Therefore, the evolution of olive cultivated germplasm in Greece may have been influenced by human-mediated gene flow of improved clones, as has been hypothesized by prior studies (Nikoloudakis et al., 2003; Owen et al., 2005; Roubos et al., 2010). Additionally, two cases of genetic similarity were noted concerning cultivars from Ionian islands with Epirus wilds, and cultivars from Peloponnese-Crete with wild samples from the same regions. For the Chania (Crete) region where wild sampling for the present study was conducted, Aksehirli-Pakyurek et al. (2017) identified genetic relatedness between wild samples and the "Mastoidis" cultivar. Taken together, these findings point to genetic pollution from historical human agricultural activity which has most likely eroded the wild gene pool. However, evidence from organelle DNA revealed that the Peloponnese area could be a contact spot between the two known wild olive gene pools of West and East Mediterranean (Besnard et al., 2013a; 2013b). Hence, the spatial structure that wild populations exhibited could also be attributed up to some extent to a phylogeographic signal.

Although a signal of genetic structure was detected among the wild olive populations, the rates of correct assignment were low. This means that assigning samples of unknown origin to geographical regions is not expected to have high reliability. Rates of correct assignment are expected to improve with an increase in sample size. Benestan et al. (2015, 2016) recorded 10,156 SNPs in lobster populations and showed that the success of assignment was substantially affected by the intensity of sampling. Specifically, for sample sizes <50, the success of assignment was limited to approximately 60%, slightly higher than the 50% expected if assignment was randomly performed. Conversely, for population sizes ≥100, successful assignment rates higher than 80% were recorded. Although collecting a large number of individuals (≥100) seems to offer high reliability in genotype assignment analyses, increasing the sample size is not always feasible, particularly in populations that have undergone bottlenecks and/or genetic pollution, as is the case for wild olive trees. Assignment rates could also be improved without increasing the sampling effort in the number of samples or genetic markers, by including morphological data. Rellstab et al. (2016) characterized almost 1,400 samples of phylogenetically close oak species in Switzerland using eight microsatellite markers and 13 morphological traits. The highest successful assignment rate was achieved by combining molecular and morphological data.

4.5 Gene flow

By investigating the data set using the software TreeMix, three migration events emerged. The first one indicated gene flow from an ancestral population of *O. europaea* subsp. *Cuspidata* to the populations of north-eastern Greece (NE-GR). Hybridization with other subspecies has been found in multiple studies, for

example, the Dhokar cultivar shares genetic origins with subspecies *laperrinei* (Besnard et al., 2013b). Genetic affinity between populations of *O. europaea* subsp. *Cuspidata* in Egypt and wild olive *O. europaea* var. *sylvestris* in the Eastern Mediterranean has also been reported (Besnard et al., 2007). However, the exchange of genetic material between the two subspecies appears to be limited to ancestral hybridization events before the creation of the natural barrier of the Sahara desert (Besnard et al., 2012; 2013b). Thus, the observed genetic link might represent an ancient event which took place in the Eastern - South-eastern Mediterranean or more likely a recent event connected to cultivar dispersion beyond the Mediterranean (Besnard et al., 2018).

The next two events involved bidirectional gene flow between the population of wild olives in Peloponnese-Crete (PL and CR) and the cultivar group C2, which was distinguished based on the clustering results and contained almost exclusively cultivars from Peloponnese, Crete and Epirus. Gene flow intensity was higher in the direction from wild populations to cultivated individuals, which possibly reveals the existence of secondary domestication events of limited geographical scope in Greece. Nevertheless, the bidirectional aspect of this gene exchange between the two groups also indicates wild population genetic erosion.

4.6 Implications for olive tree genetic resources management in Greece

The present study demonstrates that Greek wild olive tree populations withhold high levels of genetic diversity, a large proportion of which is not present in the cultivated germplasm. This is of paramount importance, regardless of the degree to which current wild individuals resemble their wild ancestors and contemporary cultivars. Although the presence of high genetic diversity in wild populations indicates that they are not at risk of immediate collapse, caution should be exerted to the fact that their ecological habitat is fragmented and surrounded by olive groves. Therefore, unless conservation measures are taken, the genetic diversity and identity of future wild olive tree generations are expected to be eroded to a higher extent than what has already occurred. The establishment of *in situ* protected areas is proposed for the conservation of isolated populations. Candidate areas have been proposed in Martsalo Gorge (Asterousia Mountains, Crete) and Mount Mandalo south of Rethymno (Banilas et al., 2009). It is expected that in such areas wild populations will be able to continue adapting to future environmental changes. For the vast majority of Greek wild populations though, geographical isolation from olive groves is not realistic. Hence the establishment of *ex situ* conservation units away from olive grove pollen flow is the only solution to maintain their genetic integrity in the long term. The selection of individuals for such a collection could be aided by the use of molecular tools, such as those used in this study, to collect plant genetic material from all genetic groups found in the country. A special case is that of centennial trees found in traditional olive groves. Many of them were propagated through grafting onto wild rootstocks, which means that they retain at their base a snapshot of the genetic diversity that existed in each region hundreds of years ago. Recording and protection of these olive groves is of the utmost importance. Finally, in a rather limited set of 14 cultivar samples

from olive groves, three of them (all from Lesvos island) were not matched to any of the reference samples from the Greek National Germplasm Bank. Growers characterized these unidentified trees as local cultivars, which underscores the need for thorough recording and conservation of the country's cultivated germplasm.

4.7 Utilization of wild olive trees in breeding, use as rootstocks

The domestication process from which olive cultivars have emerged is associated with genetic bottlenecks due to selection from local gene pools (Besnard et al., 2013a; Diez et al., 2015). Evidence from the present study as well as from prior works (please add references) suggest that wild olive tree populations exhibit unique genetic polymorphisms that are not found in the cultivated genetic material (Barazani et al., 2014; 2016; Mariotti et al., 2020; Zunino et al., 2023). Therefore, they are expected to constitute an invaluable source of variation for agronomic traits of interest. With regard to olive oil quality, Leon et al. (2018) studied the chemical composition of olive oil of cultivars, wild individuals and their hybrids. They concluded that the utilization of wild material did not have negative effects on fatty acid composition, tocopherol content, and tocopherol and phytosterol profiles if selection for these traits was carried out in the early generations of the breeding programs. In fact, it was found that tocopherol content could be improved by using wild parents. In another wild olive progeny test, higher vigour, a shorter juvenile phase and denser flowering were recorded compared to open pollination families of the cultivar "Picual" (Klepo et al., 2013). Interestingly, high resistance of wild olive genotypes to verticillium has been recorded, which in some cases was higher compared to the resistant cultivars (Colella et al., 2008; León et al., 2014). Wild olive trees are also regularly used as rootstocks with high resistance to biotic/abiotic factors as explained above. Research in rootstock management is also a very promising field of study, as there is evidence of non-random clustering of rootstocks and cultivars (Barazani et al., 2017). Therefore, the identification of unexploited adaptive traits in wild olive genetic resources and their potential utilization as rootstocks for olive cultivation, notably in Mediterranean-climate regions, could represent a major goal of olive crop breeding.

4.8 Limitations

Here we demonstrate that the 10 nuclear microsatellites analysed in this study were highly informative in discriminating olive tree accessions and elucidating patterns of genetic structure and gene flow. Nevertheless, using a high number of genomic markers in samples of a greater geographical area in the future would increase the robustness of such a study. In fact, a relatively low genetic differentiation was recorded between wild and cultivated trees, as well as among wild populations. The precision of this rather weak signal of fine genetic structure might have been lessened by the employment of a limited number of markers (Wang, 2017). Genotyping of organelle genetic markers and especially the use of high throughput genome sequencing methods are expected to provide higher resolution to estimate trends of effective population sizes (Julca et al., 2020), accurately determine hybridization events among wild and

cultivated olive tree gene pools (Bazakos et al., 2023), and elucidate olive tree evolutionary history (Julca et al., 2023). Moreover, as inferences regarding gene flow/minor domestication events are based on wild olive/cultivar reference samples that mostly originated within Greece, they should ideally be further validated with an extended data set.

5 Conclusion

The contemporary wild olive gene pool is unique, rich and unexplored. It is the result of the interaction of complex natural phylogenetic processes and anthropogenic influence through the indirect reinforcement of hybridization by the spread of improved germplasm. In the present study, molecular genetic analysis uncovered a total of 86 alleles (40% of the total number of alleles detected) which were not present in the Greek-origin cultivars. This fact demonstrates the existence of a diverse genetic resource which cannot possibly have originated purely from hybridisations between cultivars and therefore needs to be protected both *in situ* and *ex situ*. Special attention should also be given to centennial olive trees residing within traditional olive groves. As grafting with wild rootstocks was a highly common practice for grove establishment in Greece, these trees act as preservation vessels for potentially lost wild populations and should be explored and protected. From the perspective of cultivar breeding, the high allelic richness recorded is indicative of the potential of wild genetic material to be used in cultivar improvement and rootstocks. Genetic differentiation between wild individuals and cultivars was low and only weak genetic structure was observed between the wild populations. In addition, results from bayesian and gene flow analysis revealed signs of genetic pollution for populations from western Greece. Taken together, wild samples analysed in this work are probably best described as feral, with western Greek populations having endured higher genetic erosion than the rest which might resemble closer the genetic composition of their wild progenitors. Finally, the genetic profile matching of "Throuba Naxou" with "Throubolia" cultivar along with the uncovering of a gene exchange event from wild to cultivated gene pool are indications which support the hypothesized role of the Greek region as a hotspot of oleiculture evolution in the form of secondary domestication events.

Data availability statement

The datasets generated during the current study are available in the Zenodo repository, <https://doi.org/10.5281/zenodo.8104610>. Analysis code and a fully reproducible computational environment is available in the same repository.

Author contributions

NT: Data curation, Investigation, Methodology, Validation, Formal Analysis, Resources, Visualization, Writing—original draft. IG: Data curation, Investigation, Methodology, Resources, Validation, Conceptualization, Writing—review and editing. GKou: Investigation, Resources, Validation, Writing—review and editing. GKos: Resources, Validation, Writing—review and editing. IM: Investigation, Validation, Writing—review and editing. CB: Validation, Writing—review and

editing. VS: Investigation, Validation, Writing–review and editing. AM: Conceptualization, Funding acquisition, Project administration, Validation, Writing–review and editing. FA: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research has been financed by Greek national funds through the Public Investments Program (PIP) of General Secretariat for Research and Technology (GSRT), under the Emblematic Action “The Olive Road” (project code: 2018ΣΕ01300000) and by project Crown Genome, funded by the General Secretariat for Research and Innovation, Greece and the Aristotle University of Thessaloniki (project code 40220).

Acknowledgments

Results presented in this work have been produced using the Aristotle University of Thessaloniki (AUTh) High Performance Computing Infrastructure and Resources.

References

- Akshirli-Pakyurek, M., Koubouris, G. C., Petrakis, P. V., Hepaksoy, S., Metzidakis, I. T., Yalcinkaya, E., et al. (2017). Cultivated and wild olives in Crete, Greece—genetic diversity and relationships with major Turkish cultivars revealed by SSR markers. *Plant Mol. Biol. Rep.* 35, 575–585. doi:10.1007/s11105-017-1046-y
- Alcántara, J. M., and Rey, P. J. (2003). Conflicting selection pressures on seed size: evolutionary ecology of fruit size in a bird-dispersed tree, *Olea europaea*. *J. Evol. Biol.* 16, 1168–1176. doi:10.1046/j.1420-9101.2003.00618.x
- Anderson, E. C. (2010). Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. *Mol. Ecol. Resour.* 10, 701–710. doi:10.1111/j.1755-0998.2010.02846.x
- Archer, F. I., Adams, P. E., and Schneiders, B. B. (2017). Stratag: an R package for manipulating, summarizing and analysing population genetic data. *Mol. Ecol. Resour.* 17, 5–11. doi:10.1111/1755-0998.12559
- Arnaud-Haond, S., Duarte, C. M., Alberto, F., and Serrão, E. A. (2007). Standardizing methods to address clonality in population studies. *Mol. Ecol.* 16, 5115–5139. doi:10.1111/j.1365-294X.2007.03535.x
- Baldoni, L., Cultrera, N. G., Mariotti, R., Ricciolini, C., Arcioni, S., Vendramin, G. G., et al. (2009). A consensus list of microsatellite markers for olive genotyping. *Mol. Breed.* 24, 213–231. doi:10.1007/s11032-009-9285-8
- Baldoni, L., Tosti, N., Ricciolini, C., Belaj, A., Arcioni, S., Pannelli, G., et al. (2006). Genetic structure of wild and cultivated olives in the central Mediterranean basin. *Ann. Bot.* 98, 935–942. doi:10.1093/aob/mcl178
- Banilas, G., Korkas, E., Kaldis, P., and Hatzopoulos, P. (2009). “Olive and grapevine biodiversity in Greece and Cyprus – a review,” in *Climate change, intercropping, pest control and beneficial microorganisms: Climate change, intercropping, pest control and beneficial microorganisms sustainable agriculture reviews*. Editor E. Lichtfouse (Dordrecht, Netherlands: Springer), 401–428. doi:10.1007/978-90-481-2716-0_14
- Barazani, O., Keren-Keiserman, A., Westberg, E., Hanin, N., Dag, A., Ben-Ari, G., et al. (2016). Genetic variation of naturally growing olive trees in Israel: from abandoned groves to feral and wild? *BMC Plant Biol.* 16, 261. doi:10.1186/s12870-016-0947-5
- Barazani, O., Waitz, Y., Tugendhaft, Y., Dorman, M., Dag, A., Hamidat, M., et al. (2017). Testing the potential significance of different scion/rootstock genotype combinations on the ecology of old cultivated olive trees in the southeast Mediterranean area. *BMC Ecol.* 17, 3. doi:10.1186/s12898-017-0114-3
- Barazani, O., Westberg, E., Hanin, N., Dag, A., Kerem, Z., Tugendhaft, Y., et al. (2014). A comparative analysis of genetic variation in rootstocks and scions of old olive trees – a window into the history of olive cultivation practices and past genetic variation. *BMC Plant Biol.* 14, 146. doi:10.1186/1471-2229-14-146
- Barranco, D., and Rallo, L. (2000). Olive cultivars in Spain. *HortTechnology* 10, 107–110. doi:10.21273/HORTTECH.10.1.107
- Bartolini, G., and Cerreti, S. (2008). Olive germplasm (*Olea europaea* L.), cultivars, synonyms, cultivation area, collections, descriptors. <http://www.oleadb.it/>.
- Bazakos, C., Alexiou, K. G., Ramos-Onsins, S., Koubouris, G., Tourvas, N., Xanthopoulou, A., et al. (2023). Whole genome scanning of a Mediterranean basin hotspot collection provides new insights into olive tree biodiversity and biology. *Plant J. n/a* 116, 303–319. doi:10.1111/tpj.16270
- Belaj, A., Muñoz-Diez, C., Baldoni, L., Porceddu, A., Barranco, D., and Satovic, Z. (2007). Genetic diversity and population structure of wild olives from the north-western Mediterranean assessed by SSR markers. *Ann. Bot.* 100, 449–458. doi:10.1093/aob/mcm132
- Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., and Bernatchez, L. (2015). RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Mol. Ecol.* 24, 3299–3315. doi:10.1111/mec.13245
- Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., and Bernatchez, L. (2016). Erratum- Benestan et al. 2015. *Mol. Ecol.* 25, 1626–1629. doi:10.1111/mec.13600
- Besnard, G., Anthelme, F., and Baali-Cherif, D. (2012). The Laperrine’s olive tree (*Oleaaceae*): a wild genetic resource of the cultivated olive and a model-species for studying the biogeography of the Saharan Mountains. *Acta Bot. Gallica* 159, 319–328. doi:10.1080/12538078.2012.724281
- Besnard, G., Casas, R. R. de, and Vargas, P. (2007). Plastid and nuclear DNA polymorphism reveals historical processes of isolation and reticulation in the olive tree complex (*Olea europaea*). *J. Biogeogr.* 34, 736–752. doi:10.1111/j.1365-2699.2006.01653.x
- Besnard, G., El Bakkali, A., Haouane, H., Baali-Cherif, D., Moukhli, A., and Khadari, B. (2013a). Population genetics of Mediterranean and Saharan olives: geographic patterns of differentiation and evidence for early generations of admixture. *Ann. Bot.* 112, 1293–1302. doi:10.1093/aob/mct196
- Besnard, G., Khadari, B., Navascués, M., Fernández-Mazuecos, M., Bakkali, A. E., Arrigo, N., et al. (2013b). The complex history of the olive tree: from Late Quaternary diversification of Mediterranean lineages to primary domestication in the northern Levant. *Proc. R. Soc. B* 280, 20122833. doi:10.1098/rspb.2012.2833
- Besnard, G., Khadari, B., Villemur, P., and Bervillé, A. (2000). Cytoplasmic male sterility in the olive (*Olea europaea* L.). *Theor. Appl. Genet.* 100, 1018–1024. doi:10.1007/s001220051383

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2023.1298565/full#supplementary-material>

- Besnard, G., Terral, J.-F., and Cornille, A. (2018). On the origins and domestication of the olive: a review and perspectives. *Ann. Bot.* 121, 385–403. doi:10.1093/aob/mcx145
- Bracci, T., Sebastiani, L., Busconi, M., Fogher, C., Belaj, A., and Trujillo, I. (2009). SSR markers reveal the uniqueness of olive cultivars from the Italian region of Liguria. *Sci. Hortic.* 122, 209–215. doi:10.1016/j.scienta.2009.04.010
- Breton, C. M., Farinelli, D., Koubouris, G., Famiani, F., Raymond, M., and Bervillé, A. (2021). A dual-successive-screen model at pollen/stigma and pollen tube/ovary explaining paradoxical self-incompatibility diagnosis in the olive tree—an interpretative update of the literature. *Plants* 10, 1938. doi:10.3390/plants10091938
- Breton, C. M., Terral, J.-F., Pinatel, C., Médail, F., Bonhomme, F., and Bervillé, A. (2009). The origins of the domestication of the olive tree. *C. R. Biol.* 332, 1059–1064. doi:10.1016/j.crv.2009.08.001
- Breton, C. M., Tersac, M., and Bervillé, A. (2006). Genetic diversity and gene flow between the wild olive (oleaster, *Olea europaea* L.) and the olive: several Plio-Pleistocene refuge zones in the Mediterranean basin suggested by simple sequence repeats analysis. *J. Biogeogr.* 33, 1916–1928. doi:10.1111/j.1365-2699.2006.01544.x
- Carriero, F., Fontanazza, G., Cellini, F., and Giorio, G. (2002). Identification of simple sequence repeats (SSRs) in olive (*Olea europaea* L.). *Theor. Appl. Genet.* 104, 301–307. doi:10.1007/s001220100691
- Carrión, Y., Ntinou, M., and Badal, E. (2010). *Olea europaea* L. In the north Mediterranean basin during the pleniglacial and the early-middle holocene. *Quat. Sci. Rev.* 29, 952–968. doi:10.1016/j.quascirev.2009.12.015
- Chen, K.-Y., Marschall, E. A., Sovic, M. G., Fries, A. C., Gibbs, H. L., and Ludsins, S. A. (2018). assign<sc>POP</sc>: an R package for population assignment using genetic, non-genetic, or integrated data in a machine-learning framework. *Methods Ecol. Evol.* 9, 439–446. doi:10.1111/2041-210X.12897
- Cipriani, G., Marrazzo, M. T., Di Gasparo, G., Pfeiffer, A., Morgante, M., and Testolin, R. (2008). A set of microsatellite markers with long core repeat optimized for grape (*Vitis* spp.) genotyping. *BMC Plant Biol.* 8, 127. doi:10.1186/1471-2229-8-127
- Cipriani, G., Marrazzo, M. T., Marconi, R., Cimato, A., and Testolin, R. (2002). Microsatellite markers isolated in olive (*Olea europaea* L.) are suitable for individual fingerprinting and reveal polymorphism within ancient cultivars. *Theor. Appl. Genet.* 104, 223–228. doi:10.1007/s001220100685
- Clark, L. V., and Jasieniuk, M. (2011). polysat: an R package for polyploid microsatellite analysis. *Mol. Ecol. Resour.* 11, 562–566. doi:10.1111/j.1755-0998.2011.02985.x
- Colella, C., Miacola, C., Amenduni, M., D'Amico, M., Bubici, G., and Cirulli, M. (2008). Sources of verticillium wilt resistance in wild olive germplasm from the Mediterranean region. *Plant Pathol.* 57, 533–539. doi:10.1111/j.1365-3059.2007.01785.x
- Contento, A., Ceccarelli, M., Gelati, M. T., Maggini, F., Baldoni, L., and Cionini, P. G. (2002). Diversity of *Olea* genotypes and the origin of cultivated olives. *Theor. Appl. Genet.* 104, 1229–1238. doi:10.1007/s00122-001-0799-7
- Dewoody, J., Nason, J. D., and Hipkins, V. D. (2006). Mitigating scoring errors in microsatellite data from wild populations: REVIEW. *Mol. Ecol. Notes* 6, 951–957. doi:10.1111/j.1471-8286.2006.01449.x
- Díaz, A., De la Rosa, R., Martín, A., and Rallo, P. (2006). Development, characterization and inheritance of new microsatellites in olive (*Olea europaea* L.) and evaluation of their usefulness in cultivar identification and genetic relationship studies. *Tree Genet. Genomes* 2, 165–175. doi:10.1007/s11295-006-0041-5
- Díez, C. M., Trujillo, I., Barrio, E., Belaj, A., Barranco, D., and Rallo, L. (2011). Centennial olive trees as a reservoir of genetic diversity. *Ann. Bot.* 108, 797–807. doi:10.1093/aob/mcr194
- Díez, C. M., Trujillo, I., Martínez-Urdiroz, N., Barranco, D., Rallo, L., Marfil, P., et al. (2015). Olive domestication and diversification in the Mediterranean basin. *New Phytol.* 206, 436–447. doi:10.1111/nph.13181
- Doveri, S., Sabino Gil, F., Díaz, A., Reale, S., Busconi, M., da Câmara Machado, A., et al. (2008). Standardization of a set of microsatellite markers for use in cultivar identification studies in olive (*Olea europaea* L.). *Sci. Hortic.* 116, 367–373. doi:10.1016/j.scienta.2008.02.005
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x
- Falush, D., Stephens, M., and Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164, 1567–1587. doi:10.1093/genetics/164.4.1567
- Fitak, R. R. (2021). OptM: estimating the optimal number of migration edges on population trees using Treemix. *Biol. Methods Protoc.* 6, bpab017. doi:10.1093/biomethods/bpab017
- Flores-Rentería, L., and Krohn, A. (2013). “Scoring microsatellite loci,” in *Microsatellites methods in molecular biology*. Editor S. K. Kantartzis (Totowa, NJ, USA: Humana Press), 319–336. doi:10.1007/978-1-62703-389-3_21
- Francis, R. M. (2017). pophler: an R package and web app to analyse and visualize population structure. *Mol. Ecol. Resour.* 17, 27–32. doi:10.1111/1755-0998.12509
- Gemas, V. J. V., Almadanim, M. C., Tenreiro, R., Martins, A., and Fevereço, P. (2004). Genetic diversity in the Olive tree (*Olea europaea* L. subsp. *europaea*) cultivated in Portugal revealed by RAPD and ISSR markers. *Genet. Resour. Crop Evol.* 51, 501–511. doi:10.1023/B:GRES.0000024152.16021.40
- Gilbert, K. J., Andrew, R. L., Bock, D. G., Franklin, M. T., Kane, N. C., Moore, J.-S., et al. (2012). Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program structure. *Mol. Ecol.* 21, 4925–4930. doi:10.1111/j.1365-294X.2012.05754.x
- 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
- Green, P. S. (2002). A revision of *Olea* L. (*Oleaceae*). *Kew Bull.* 57, 91. doi:10.2307/4110824
- Gros-Balthazard, M., Besnard, G., Sarah, G., Holtz, Y., Leclercq, J., Santoni, S., et al. (2012). Evolutionary transcriptomics reveals the origins of olives and the genomic changes associated with their domestication. *Plant J.* 100, 143–157. doi:10.1111/tjp.14435
- Hannachi, H., Breton, C., Msallem, M., Ben El Hadj, S., El Gazzah, M., and Bervillé, A. (2008). Differences between native and introduced olive cultivars as revealed by morphology of drupes, oil composition and SSR polymorphisms: a case study in Tunisia. *Sci. Hortic.* 116, 280–290. doi:10.1016/j.scienta.2008.01.004
- Hauwane, H., El Bakkali, A., Moukhi, A., Tollon, C., Santoni, S., Oukabli, A., et al. (2011). Genetic structure and core collection of olives and the World Olive Germplasm Bank of Marrakech: towards the optimised management and use of Mediterranean olive genetic resources. *Genetica* 139, 1083–1094. doi:10.1007/s10709-011-9608-7
- Hoffman, J. I., and Amos, W. (2005). Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Mol. Ecol.* 14, 599–612. doi:10.1111/j.1365-294X.2004.02419.x
- Jiménez-Ruiz, J., Ramírez-Tejero, J. A., Fernández-Pozo, N., de la O Leyva-Pérez, M., Yan, H., de la Rosa, R., et al. (2020). Transposon activation is a major driver in the genome evolution of cultivated olive trees (*Olea europaea* L.). *Plant Genome* 13, e20010. doi:10.1002/tpg2.20010
- Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405. doi:10.1093/bioinformatics/btn129
- Jost, L. (2008). G_{ST} and its relatives do not measure differentiation. *Mol. Ecol.* 17, 4015–4026. doi:10.1111/j.1365-294X.2008.03887.x
- Julca, I., Marcet-Houben, M., Cruz, F., Gómez-Garrido, J., Gaut, B. S., Díez, C. M., et al. (2020). Genomic evidence for recurrent genetic admixture during the domestication of Mediterranean olive trees (*Olea europaea* L.). *BMC Biol.* 18, 148. doi:10.1186/s12915-020-00881-6
- Julca, I., Vargas, P., and Gabaldón, T. (2023). Phylogenomics of the *Olea europaea* complex using 15 whole genomes supports recurrent genetic admixture together with differentiation into seven subspecies. *BMC Biol.* 21, 85. doi:10.1186/s12915-023-01583-5
- Kamvar, Z. N., Tabima, J. F., and Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2, e281. doi:10.7717/peerj.281
- Kaniewski, D., Van Campo, E., Boiy, T., Terral, J.-F., Khadari, B., and Besnard, G. (2012). Primary domestication and early uses of the emblematic olive tree: palaeobotanical, historical and molecular evidence from the Middle East. *Biol. Rev.* 87, 885–899. doi:10.1111/j.1469-185X.2012.00229.x
- Khadari, B., Breton, C., Moutier, N., Roger, J., Besnard, G., Bervillé, A., et al. (2003). The use of molecular markers for germplasm management in a French olive collection. *Theor. Appl. Genet.* 106, 521–529. doi:10.1007/s00122-002-1079-x
- Khadari, B., and El Bakkali, A. (2018). Primary selection and secondary diversification: two key processes in the history of olive domestication. *Int. J. Agron.* 2018, 1–9. doi:10.1155/2018/5607903
- Klepo, T., De la Rosa, R., Satovic, Z., León, L., and Belaj, A. (2013). Utility of wild germplasm in olive breeding. *Sci. Hortic.* 152, 92–101. doi:10.1016/j.scienta.2012.12.010
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* 15, 1179–1191. doi:10.1111/1755-0998.12387
- Kostelenos, G., and Kiritsakis, A. (2017). “Olive tree history and evolution,” in *Olives and olive oil as functional foods* (Hoboken, New Jersey, United States: John Wiley and Sons, Ltd), 1–12. doi:10.1002/9781119135340.ch1
- Koubouris, G. C., Avramidou, E. V., Metzdakis, I. T., Petrakis, P. V., Sergentani, C. K., and Doulis, A. G. (2019). Phylogenetic and evolutionary applications of analyzing endocarp morphological characters by classification binary tree and leaves by SSR markers for the characterization of olive germplasm. *Tree Genet. Genomes* 15, 26. doi:10.1007/s11295-019-1322-0
- Koubouris, G. C., Breton, C. M., Metzdakis, I. T., and Vasilakakis, M. D. (2014). Self-incompatibility and pollination relationships for four Greek olive cultivars. *Sci. Hortic.* 176, 91–96. doi:10.1016/j.scienta.2014.06.043

- León, L., Arias-Calderón, R., Bejarano-Alcázar, J., De La Rosa, R., Belaj, A., and Rodríguez-Jurado, D. (2014). Screening of wild olives for Verticillium wilt resistance. *Acta Hort.* 1057, 559–563. doi:10.17660/ActaHortic.2014.1057.71
- León, L., de la Rosa, R., Velasco, L., and Belaj, A. (2018). Using wild olives in breeding programs: implications on oil quality composition. *Front. Plant Sci.* 9, 232. doi:10.3389/fpls.2018.00232
- Lumaret, R., Ouazzani, N., Michaud, H., Vivier, G., Deguilloux, M.-F., and Di Giusto, F. (2004). Allozyme variation of oleaster populations (wild olive tree) (*Olea europaea* L.) in the Mediterranean Basin. *Heredity* 92, 343–351. doi:10.1038/sj.hdy.6800430
- Mariotti, R., Belaj, A., De La Rosa, R., León, L., Brizioli, F., Baldoni, L., et al. (2020). EST-SNP study of *Olea europaea* L. Uncovers functional polymorphisms between cultivated and wild olives. *Genes* 11, 916. doi:10.3390/genes11080916
- Martínez, S. R., and Gandullo, J. M. (1987). *Mapa de series de vegetación de España y Memoria: 1: 400.000*. Icona. Available at: https://www.miteco.gob.es/es/biodiversidad/servicios/banco-datos-naturaleza/informacion-disponible/memoria_mapa_series_veg.html.
- Marwick, B., Boettiger, C., and Mullen, L. (2018). Packaging data analytical work reproducibly using R (and friends). *Am. Stat.* 72, 80–88. doi:10.1080/00031305.2017.1375986
- Meirmans, P. G., and Hedrick, P. W. (2011). Assessing population structure: F(ST) and related measures. *Mol. Ecol. Resour.* 11, 5–18. doi:10.1111/j.1755-0998.2010.02927.x
- Mulas, M., and Deidda, P. (1998). Domestication of woody plants from Mediterranean maquis to promote new crops for mountain lands. *Acta Hort.* 457, 295–302. doi:10.17660/actahortic.1998.457.37
- Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89, 583–590. doi:10.1093/genetics/89.3.583
- Neophytou, C. (2014). Bayesian clustering analyses for genetic assignment and study of hybridization in oaks: effects of asymmetric phylogenies and asymmetric sampling schemes. *Tree Genet. Genomes* 10, 273–285. doi:10.1007/s11295-013-0680-2
- Nikoloudakis, N., Banilas, G., Gazis, F., Hatzopoulos, P., and Metzidakis, J. (2003). Discrimination and genetic diversity among cultivated olives of Greece using RAPD markers. *J. Am. Soc. Hortic. Sci.* 128, 741–746. doi:10.21273/JASHS.128.5.0741
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2020). *vegan: community ecology package*. <https://github.com/vegandevs/vegan>.
- Owen, C. A., Bitá, E.-C., Banilas, G., Hajjar, S. E., Sellianakis, V., Aksoy, U., et al. (2005). AFLP reveals structural details of genetic diversity within cultivated olive germplasm from the Eastern Mediterranean. *Theor. Appl. Genet.* 110, 1169–1176. doi:10.1007/s00122-004-1861-z
- Paetkau, D., and Strobeck, C. (1994). Microsatellite analysis of genetic variation in black bear populations. *Mol. Ecol.* 3, 489–495. doi:10.1111/j.1365-294X.1994.tb00127.x
- 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
- Pina-Martins, F., Silva, D. N., Fino, J., and Paulo, O. S. (2017). *Structure_threader*: an improved method for automation and parallelization of programs structure, *fastStructure* and *MavericK* on multicore CPU systems. *Mol. Ecol. Resour.* 17, e268–e274. doi:10.1111/1755-0998.12702
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi:10.1093/genetics/155.2.945
- Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Mol. Ecol. Resour.* 16, 608–627. doi:10.1111/1755-0998.12512
- R Core Team (2021). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: <https://www.r-project.org/>.
- Relstab, C., Buhler, A., Graf, R., Folly, C., and Gugerli, F. (2016). Using joint multivariate analyses of leaf morphology and molecular-genetic markers for taxon identification in three hybridizing European white oak species (*Quercus* spp.). *Ann. For. Sci.* 73, 669–679. doi:10.1007/s13595-016-0552-7
- Roubos, K., Moustakas, M., and Aravanopoulos, F. A. (2010). Molecular identification of Greek olive (*Olea europaea*) cultivars based on microsatellite loci. *Genet. Mol. Res.* 9, 1865–1876. doi:10.4238/vol9-3gmr916
- Rousset, F. (2008). *genepop'007*: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* 8, 103–106. doi:10.1111/j.1471-8286.2007.01931.x
- Rubio, R., Balaguer, L., Manrique, E., Pérez, M. E., and Vargas, P. (2002). "On the historical presence of the wild olive [*Olea europaea* L. var. *sylvestris* (Miller): *lehr*. (*Oleaceae*)] in the Eurosiberian region of the Iberian Peninsula," in *Anales del Jardín Botánico de Madrid* (Chamartín: Consejo Superior de Investigaciones Científicas), 342–344.
- Sefc, K. M., Lopes, M. S., Mendonça, D., Santos, M. R. D., Machado, M. L. D. C., and Machado, A. D. C. (2000). Identification of microsatellite loci in olive (*Olea europaea*) and their characterization in Italian and Iberian olive trees. *Mol. Ecol.* 9, 1171–1173. doi:10.1046/j.1365-294X.2000.00954.x
- Taberlet, P., and Luikart, G. (1999). Non-invasive genetic sampling and individual identification. *Biol. J. Linn. Soc.* 68, 41–55. doi:10.1111/j.1095-8312.1999.tb01157.x
- Terral, J.-F., Alonso, N., Buxó I Capdevila, R., Chatti, N., Fabre, L., Fiorentino, G., et al. (2004). Historical biogeography of olive domestication (*Olea europaea* L.) as revealed by geometrical morphometry applied to biological and archaeological material. *J. Biogeogr.* 31, 63–77. doi:10.1046/j.0305-0270.2003.01019.x
- Theophrastus, (1916). in *Enquiry into plants, (Greek: peri phyton historia)*. Books 1-5 (Cambridge, MA, USA: Harvard University Press).
- Trujillo, I., Ojeda, M. A., Urdiroz, N. M., Potter, D., Barranco, D., Rallo, L., et al. (2014). Identification of the worldwide olive germplasm Bank of Córdoba (Spain) using SSR and morphological markers. *Tree Genet. Genomes* 10, 141–155. doi:10.1007/s11295-013-0671-3
- Unver, T., Wu, Z., Sterck, L., Turktaş, M., Lohaus, R., Li, Z., et al. (2017). Genome of wild olive and the evolution of oil biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 114, E9413–E9422. doi:10.1073/pnas.1708621114
- Wang, J. (2017). The computer program structure for assigning individuals to populations: easy to use but easier to misuse. *Mol. Ecol. Resour.* 17, 981–990. doi:10.1111/1755-0998.12650
- Yoruk, B., and Taskin, V. (2014). Genetic diversity and relationships of wild and cultivated olives in Turkey. *Plant Syst. Evol.* 300, 1247–1258. doi:10.1007/s00606-014-1002-3
- Zunino, L., Cubry, P., Sarah, G., Mournet, P., Bakkali, A. E., Aqbouch, L., et al. (2023). Genomic evidence of genuine wild versus admixed olive populations evolving in the same natural environments in western Mediterranean Basin. <https://www.biorxiv.org/content/10.1101/2023.04.18.537296v1>.