Genetic diversity and divergence at the *Arbutus unedo* L. (Ericaceae) westernmost distribution limit

Maria Margarida Ribeiro[1,2,3]*, **Andrea Piotti**[4], **Alexandra Ricardo**[1], **Daniel Gaspar**[5], **Rita Costa**[3,5], **Laura Parducci**[2], **Giovanni Giuseppe Vendramin**[4]

1 Departamento de Recursos Naturais e Desenvolvimento Sustentável, Instituto Politécnico de Castelo Branco, Escola Superior Agrária, Castelo Branco, Portugal, **2** Plant Ecology/ Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18 D, Uppsala, Sweden, **3** Forest Research Centre, School of Agriculture, University of Lisbon, Tapada da Ajuda, Lisbon, Portugal, **4** Institute of Biosciences and BioResources, National Research Council, Via Madonna del Piano 10, Sesto Fiorentino (Firenze), Italy, **5** INIAV, Instituto Nacional de Investigação Agrária e Veterinária, I.P., Av. República, Quinta do Marquês Oeiras, Portugal

* mataide@ipcb.pt

Abstract

Mediterranean forests are fragile ecosystems vulnerable to recent global warming and reduction of precipitation, and a long-term negative effect is expected on vegetation with increasing drought and in areas burnt by fires. We investigated the spatial distribution of genetic variation of *Arbutus unedo* in the western Iberia Peninsula, using plastid markers with conservation and provenance regions design purposes. This species is currently undergoing an intense domestication process in the region, and, like other species, is increasingly under the threat from climate change, habitat fragmentation and wildfires. We sampled 451 trees from 15 natural populations from different ecological conditions spanning the whole species' distribution range in the region. We applied Bayesian analysis and identified four clusters (north, centre, south, and a single-population cluster). Hierarchical AMOVA showed higher differentiation among clusters than among populations within clusters. The relatively low within-clusters differentiation can be explained by a common postglacial history of nearby populations. The genetic structure found, supported by the few available palaeobotanical records, cannot exclude the hypothesis of two independent *A. unedo* refugia in western Iberia Peninsula during the Last Glacial Maximum. Based on the results we recommend a conservation strategy by selecting populations for conservation based on their allelic richness and diversity and careful seed transfer consistent with current species' genetic structure.

Introduction

The contemporary genetic structure of a species offers important information about its responses to past geological and climatic events, which had often played a crucial role in shaping the current distribution range. The spatial distribution of genetic variation is strongly

influenced by neutral and adaptive processes and, at the population level, the amount of extant genetic variation plays a key role in maintaining the adaptive potential to face future environmental changes [1]. Among other factors, the lack of genetic connectivity between populations can accelerate the within-population genetic loss in woody plants [2]. In the Mediterranean basin, geographical isolation at the biogeographical scale was mainly promoted by Pleistocene climatic oscillations, which determined the presence of refugial areas often hosting different genetic lineages of plant species [3]. The Iberian Peninsula (IP) is one of the major Pleistocene glacial refugial areas for European species, where multiple refugia exist [4,5]. Mediterranean forests growing on the IP are fragile ecosystems currently threatened by human-related activities like grazing, fires and intensive forest cutting. This region is also particularly vulnerable to recent global warming and reduction of precipitation, and simulation studies show a long-term negative effect on vegetation due to increase in drought and burnt area by fires [6].

The evergreen *Arbutus unedo* L. (strawberry tree) is a small tree native to the Mediterranean region found in western, central, and southern Europe, north-eastern Africa, Canary Islands and western Asia. The species distribution is mostly typical of Mediterranean sclerophyllous and laurel-like vegetation, mainly within coastal and inland areas, where frost or summer dryness are not too intense [7]. *Arbutus unedo* genetic structure has been investigated at both local and biogeographical scale using different types of markers [5,8,9,10,11,12,13]. In Portugal, strawberry tree has a widespread distribution strongly related to soil characteristics and climate, as well as to landscape fragmentation, wildfires, and domestication [14]. *Arbutus unedo* fruit has strong antioxidant properties and it can be consumed fresh or processed, and it is used in Portugal to produce the liquor 'aguardente', an important source of income for land users, which has led to an on-going process of domestication of the species through clonal propagation of genotypes selected for fruit quality [15] and to the use of seeds from unknown origin. This species has economic importance to be used in Portugal and southern Europe, with different commercial uses from processed and fresh fruit production to ornamental, pharmaceutical and chemical industrial applications, due to the phenolic acids and terpenoid compounds with strong antioxidant activity, vitamin C and tannin content [15 and references therein, 16, 17]. Nevertheless, this species remains largely underutilized, and organizations such as FAO [18] developed efforts to increase the use of those species.

Anthropogenic impacts on forest genetic resources include practices and processes that may affect the genetic layout of natural populations, including: (i) breeding and selection, (ii) seed transfer and (iii) environment alterations [19]. The effect of wildfires can greatly impact the distribution of genetic variation at regional scale, since a likely outcome of repeated wildfires is the effective population size reduction and the loss of within-population genetic diversity [20]. Regions with more severe droughts might also be affected by an increase of wildfire activity and intensity, which may have large impacts on vegetation density and distribution. In the Mediterranean basin, for instance, the area burned by wildfire is expected to increase by a factor of 3 to 5 by the end of the XXI century [21]. Nowadays, the total number of wildfires and the percentage of burned areas in Portugal are already considerably higher than in other southern Mediterranean countries, with figures showing a strong tendency to increase in the next years (S1 Fig). This is likely due to a decrease in agricultural areas in favour of woodlands [22]. Shift in land use coupled with growth of tree plantations (mainly dense coniferous stands) increased fire regularity, resulting in large and catastrophic wildfires, which in the last few decades have resulted in a marked rise of the areas burnt annually in Portugal [23].

The ongoing domestication process, together with the increasing frequency of wildfires due to higher temperatures and lower precipitations, demands further knowledge of the *A. unedo* genetic diversity at the regional scale to build dynamic conservation programs in Portugal [24 and references therein]. Due to the lack of information about the levels and distribution of

genetic diversity at the regional scale, we sampled 15 natural populations throughout the westernmost part of the IP, and genotyped 451 individuals using chloroplast microsatellite markers (cpSSRs). The chloroplast genome is maternally inherited in the strawberry tree (our unpublished data), reflecting seed flow only. Moreover, this genome has a smaller effective population size, which makes it more susceptible to genetic drift and population or species differentiation [25]. Our specific aims were to: i) investigate the spatial distribution of genetic variation of *A. unedo* in Portugal, and to ii) provide recommendations for species' conservation and help in provenance region design.

Material and methods

Study species

Arbutus unedo is an evergreen species, a shrub or a small tree up to 8–10 m. The fruit is a round berry (ca. 20 mm wide) and bright red when ripe in autumn. The berries with a fleshy pulp are mainly eaten by seed-disperser birds, which represent the main *A. unedo* seed dispersal vector [26,27]. The inflorescence is a panicle which appears in autumn, the flowers are insect pollinated [28 and our unpublished data], and crop fruit production starts when the plants are about six to seven years old (F. Gomes, personal communication). The *A. unedo* plants can easily resprout and survive if damaged (e.g., by fire or grazing) [29]. This species is a marginal woodland plant, unable to endure shading and to successfully compete for light [26]. *Arbutus unedo* prefers siliceous or decarbonated substrata and can grow on alkaline and relatively acidic soils, but grows satisfactorily under a very wide range of soil conditions, with pH varying from 4 to 7, excluding waterlogged soils [26].

Sample collection

The leaves for DNA extraction were collected in ca. 30 individuals from 15 natural populations sampled across the species' natural range in Portugal (Table 1, S2 Fig). Samples were stored at -80˚C until DNA extraction. Sampled trees were separated by a minimum distance of ca. 30 m

Table 1. Name, code, geographic location, soil and climate types of the sampled populations.

Population name	Code	Longitude[a] W	Latitude[a] N	Altitude (m a.s.l.)	Soil type	Climate type
Peneda-Gerês	PG	-8.18	41.77	611	Humic Cambisol	Tem. Oceanic
Chaves	CH	-7.44	41.71	470	Humic Cambisol	Med. Pluviseasonal Oceanic
Bragança	B	-6.95	41.5	623	Lithosol	Med. Pluviseasonal Oceanic
Gardunha	G	-7.41	40.12	728	Dystric Cambisol	Med. Pluviseasonal Oceanic
Alvoco	AV	-7.84	40.28	395	Humic Cambisol	Med. Pluviseasonal Oceanic
Oleiros Norte	ON	-7.77	39.96	678	Lithosol	Med. Pluviseasonal Oceanic
Pampilhosa da Serra	PS	-8.05	40	538	Humic Cambisol	Med. Pluviseasonal Oceanic
Sertã-Figueiredo	SF	-7.99	39.85	465	Humic Cambisol	Med. Pluviseasonal Oceanic
Arrábida	A	-9.01	38.47	252	Orthic Podzol	Med. Pluviseasonal Oceanic
Viseu	V	-7.83	40.8	641	Humic Cambisol	Med. Pluviseasonal Oceanic
São Mamede	SM	-7.41	39.4	681	Dystric Cambisol	Med. Pluviseasonal Oceanic
Herdade Parra	HP	-8.42	37.31	165	Lithosol	Med. Pluviseasonal Oceanic
Barranco do Velho	BV	-7.95	37.25	489	Lithosol	Med. Pluviseasonal Oceanic
Monchique	M	-8.52	37.33	369	Orthic Luvisol	Med. Pluviseasonal Oceanic
Espinhaço Cão	EC	-7.98	37.18	225	Chromic Luvisol	Med. Pluviseasonal Oceanic

[a] decimals degrees

to avoid collecting related individuals and the coordinates for each individual were obtained with a GPS receiver. Our sampling took into account the environmental diversity in different soil and climate types [30,31] in Portugal (Table 1).

Ethics statement

Leaf sampling of *A. unedo* adult trees was non-destructive. No specific permission was required to sample the species in Portugal. Field studies did not involve endangered or protected species.

DNA extraction, amplification and sequencing

We extracted total DNA from frozen leaves with a Dneasy® Plant Mini kit (QIAGEN, Hilden, Germany), following the manufacturer's instructions. Six chloroplast microsatellite primers (Table 2) were designed based on the chloroplast genome [32]. The genome was scanned for cpSSRs loci using the software SciRoKo 3.4 [33]. Primers were designed in the flanking regions of the detected SSR repeats using Primer3 v. 0.4.0 [34]. The primer design parameters used were: i) primer length (we selected 18–23 bp, with 20 bp being optimal), ii) PCR product sizes (100–450 bp, by using 60°C as an optimum melting temperature of the PCR primers), and iii) GC > 40%. Each forward primer was fluorescently labelled with 6-FAM dye (AB PRISM Primer Applied Biosystems). All the individuals were genotyped with the four polymorphic markers: AU1, AU2, AU4, and AU7.

Two multiplex PCR were performed using the Type-it® Microsatellite PCR Kit (Qiagen). Forward and reverse primers for each locus were mixed at equal proportion and used in the multiplex PCR reactions. The first multiplex mix (with AU2 and AU7) was prepared in a 6 μL final reaction volume containing 1 μL of DNA (~10 ng), 3 μL of 2× Type-it mix, 1.4 μL dH2O and 0.6 μL of primer pre-mix 1 (0.1 μM each primer). The second mix (with AU1 and AU4) contained the same amount of total volume, DNA and master mix and 2 μL of primer pre-mix 2 (0.1 μM each AU1 primer and 0.2 μM each AU4 primer). The two multiplexes were amplified with the same PCR profile: an initial activation step of 5 min at 95°C, followed by 30 cycles of 30 s denaturation at 95°C, 90 s annealing at 57°C and 30 s extension at 72°C, concluding with a final extension for 30 min at 60°C. The amplified products were mixed with Hi-Di™ Formamide (Applied Biosystems) and GeneScan™ 500 LIZ™ dye size standard (Applied Biosystems) according to standard protocols and separated on an AB 3500 DNA genetic analyzer

Table 2. Code, repeat unit, size, fragment start and end and position relative to the start of the fragment of the six *A. unedo* cpSSR primers used in this study.

Code	Repeat Unit	Size (bp)	Frag. start	Frag. end	Position	Primers
AU1	(A)15	312	4164	4475	80	CCGCACTTAAAAGCCGAGTA
						GGTAAATGGTGGGTTCGTTG
AU2	(T)18	294	75426	75719	24	CCAAGCATTCCCTGAAAAGA
						GGATTTGAACCTGCGACATT
AU4	(AT)5	387	21190	21571	184	TCTCCCAATTCGACCAAGTC
						AGGGGTTGTGGATACTGCTG
AU5	(AT)5	311	51000	51310	166	TGGCCAACTCTACCAAATCC
						TCCCCCAAATGCAAATTAAA
AU6	(GA)5	151	56699	56849	45	CCCTCTTGCCAGATTTATCG
						TTTTCGGGAGTTGTTTGTCC
AU7	(TA)5	376	13771	14146	119	CCAATAGCCCAAAGAAACGA
						TGCCCTAGACCTGTCCTTTG

https://doi.org/10.1371/journal.pone.0175239.t002

(Applied Biosystems). The molecular sizes of the fragments were binned using the GeneMarker® v2.4.0 software.

Data analysis

Genetic diversity. The four cpSSR fragments analysed were combined in order to derive the chloroplast haplotype of each individual. An haplotype network, a minimum spanning tree based on the differences in number of repeats between haplotypes, was created using the `mst` function and visualized through the `plot.mst` function in the *pegas* package in R [35,36]. The number of haplotypes (N_h), the unbiased haplotypic diversity based on haplotype frequencies (H_e) [37] and the haplotypic richness A_R (number of different haplotypes found when a specific sample size is sampled from that population—in our study, the population size was fixed at 30), were computed for each population using the program CONTRIB v. 1.4 [38]. The effective number of haplotypes (A_e) and the number of private haplotypes (Ph) were calculated in GenAlEx v. 6.501 [39]. CONTRIB also provided a measure of the contribution of each population to the total diversity (CT%) and the total haplotypic richness (CTR%). In both cases, the contributions were split in a first component, due to population diversity, and a second component, due to the differentiation from the remaining populations [38]. Additionally, the average Goldstein genetic distance among individuals within each population was computed according to the stepwise mutation model (SMM) estimator D^2_{sh}, [40], as modified for fully linked markers [41]. D^2_{sh} was estimated on the original data and on a transformed dataset (ng = no gaps), where the gap between alleles were eliminated to mitigate the possible impact of indels, by in-house R scripts. The relationship between the estimates of diversity (N_h and H_e) and the genetic distances among individuals within populations D^2_{sh} and $D^{2~ng}_{sh}$ was tested with Pearson's product-moment correlation in R.

Genetic and phylogeographic structure. Bayesian analysis using a spatial clustering model implemented in BAPS 5.4 [42] was used to infer population genetic structure. The use of spatial information increases the power to correctly detect the underlying population structure [43]. The analysis was repeated 10 times for each *K* (from one to 15) to obtain the optimal number of clusters. The best partition of populations into *K* clusters was identified as the one with the highest log marginal likelihood.

The extent of the genetic structure was further explored using a non-hierarchical and hierarchical analysis of molecular variance (AMOVA) implemented in the Arlequin 3.5 software [44]. We used the infinite allele mutation model (IAM) considering the distances between haplotypes as the number of different alleles, and the stepwise mutation model (SMM) considering those distances as the number of repeat units for each locus. To perform the SMM analysis, we coded cpSSR data in a binary way, as in Heuertz *et al*. [45]. Overall, this test estimated the variance components among and within populations. The hierarchical AMOVA (based on BAPS clustering) estimated the partition of the genetic variation among clusters, among populations within clusters and within populations. The significance values were computed by 1,000 permutations.

We checked for the presence of a phylogeographic structure by comparing differentiation of unordered alleles (G_{ST}) and ordered alleles (N_{ST}) with the PermutCpSSR 1.2.1 software [46]. N_{ST} is defined as a measure of genetic differentiation among populations when the distance between haplotypes is taken into account, whereas G_{ST} is a measure of genetic differentiation that makes use only of haplotypic frequencies. The significance of the two differentiation statistics was tested with 10,000 random permutations of individuals among populations. To test if the observed N_{ST} value is larger than the G_{ST}, we counted how many permuted N_{ST} values were larger than the observed N_{ST}. Statistical significance may indicate the presence of a

phylogeographic structure [46], i.e., closely related haplotypes are more often found in the same population than would be expected by chance. Finally, isolation by distance (IBD) was tested comparing genetic and log-transformed geographic distances between population pairs, using $F_{ST}/(1-F_{ST})$ as genetic distance. The statistical significance of the Mantel test correlation coefficient was obtained from 1,000 permutations using Arlequin 3.5.

Results

Genetic diversity

We obtained 15 haplotypes after genotyping 451 individuals with four polymorphic cpSSRs (S1 Table). With only one exception (i.e. population SM), the most frequent haplotype is H10 (Figs 1A and 2). The population SM differed from the others by showing haplotype H13 at high frequency (> 0.90) (Figs 1A and 2). Five populations showed private haplotypes [AV (H8, H1), SM (H4), V (H9), SF (H14), EC (H12)].

The average haplotypic diversity within populations was 0.40 (Table 3), with populations EC and ON displaying the highest haplotypic diversity (H_e) and effective number of haplotypes (A_e). The population EC showed also the highest values for N_h (6). Other populations from the central region (G, AV and PS), had H_e values over 0.50 and $A_e > 2$. The populations
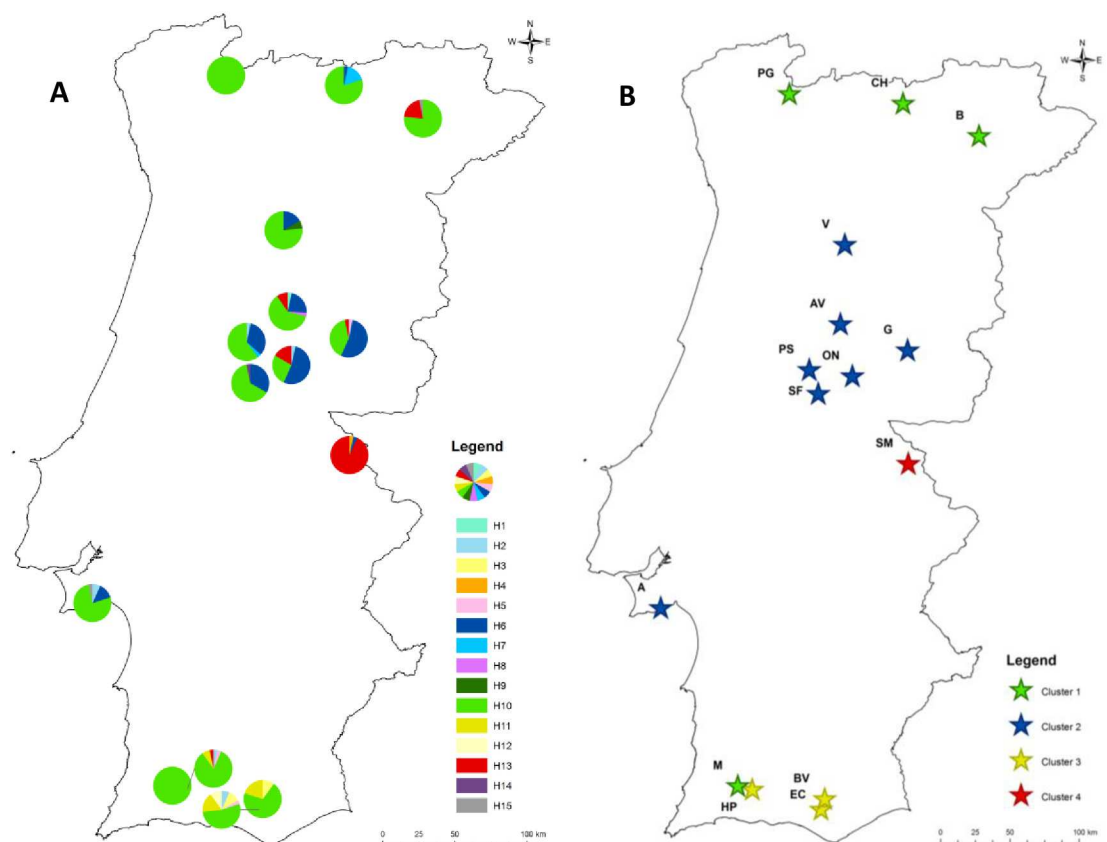
Fig 1. Haplotypes distribution and Bayesian analysis results. (A) Distribution of the haplotypes in each population. (B) Spatial clustering defined in the BAPS analysis. The best partition displayed 4 clusters: northern, southern and central clusters, and an outlier cluster, the SM population. Cluster 1: [M, PG, CH, B; red star], cluster 2: [HP, BV, EC; green star]; cluster 3: [A, SF, ON, G, AV, PS, V; blue star] and cluster 4: [SM, yellow star]. See Table 1 for population abbreviations.
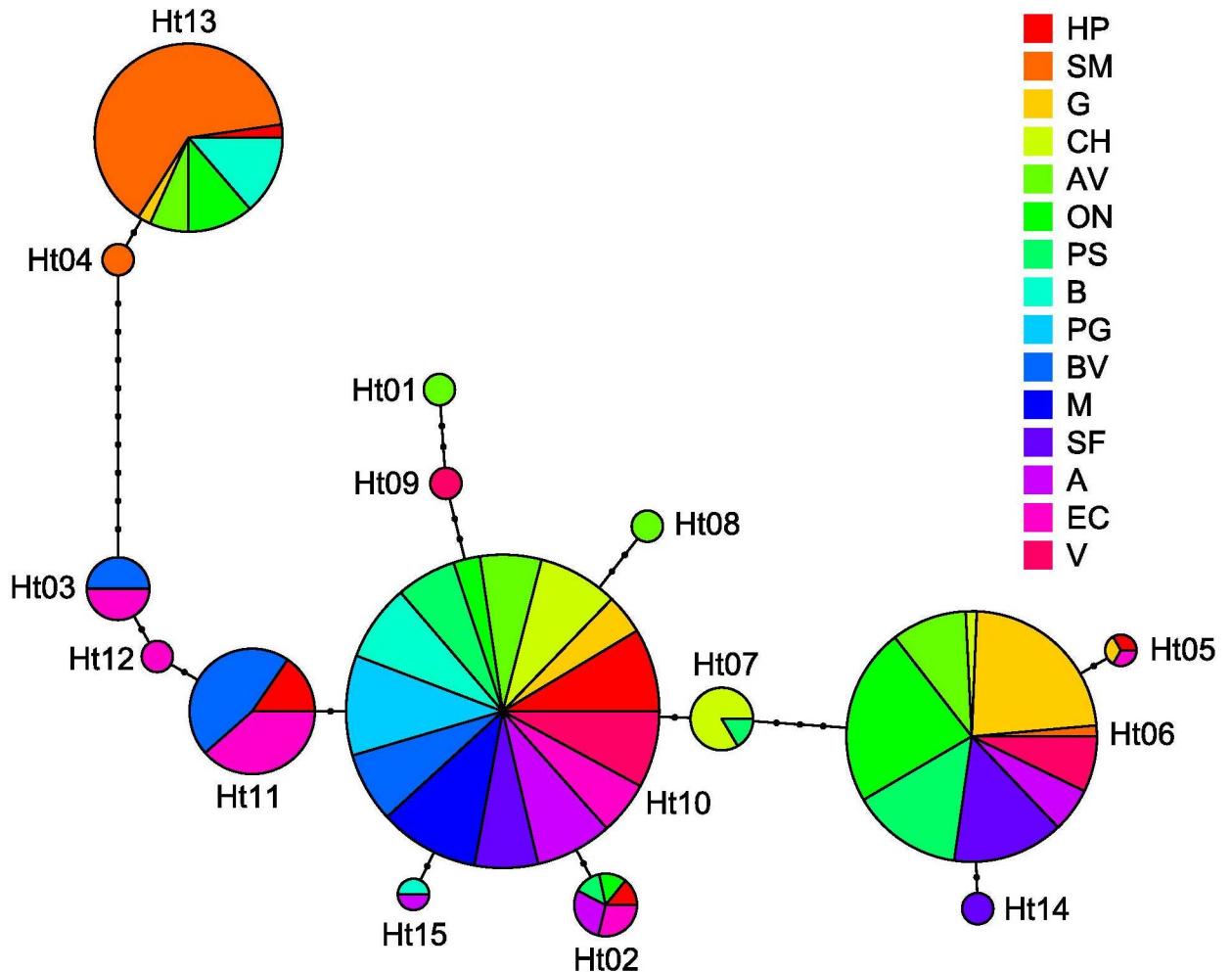
Fig 2. Haplotype network. The haplotype network is a minimum spanning tree based on the differences in number of repeats. The circles size is proportional to rescaled haplotype frequencies (see M&M for details). Colours indicate in which populations is present each haplotype. In case one haplotype is present in several populations, the circle is divided in circle sectors that are proportional to the number of plants in each population having that haplotype.

PG and M were fixed for the most common haplotype and also the population SM showed a very low value of diversity (0.13).

We found no correlation between D^2_{sh} and genetic diversity estimates (Table 3). The populations with high values of D^2_{sh}, such as the B population, did not display high values of genetic diversity (N_h and H_e). However, if we consider the genetic distance with no gaps ($D^2_{sh}{}^{ng}$), a significant correlation was found between the average genetic distance and the estimates of genetic diversity (N_h and H_e) (Table 3). This discrepancy was probably due to the presence of an insertion determining a 9 bp gap between alleles 285 and 294 at locus AU2 (see S1 Table).

The population SM showed the highest contribution to the differentiation of populations, while populations G and ON showed the highest contribution to the total diversity. The populations that gave the highest contribution to the population diversity were AV, ON and EC. The populations PG and M contributed negatively to both components of the total diversity (diversity and divergence components), showing only one haplotype (CT%, Fig 3A). The population EC had the highest contribution to the total allelic richness, followed by populations SM and BV (CTR%, Fig 3B).

Table 3. Main within-population genetic parameters for the 15 populations analysed in this study. N = population sample size; N_h = number of haplotypes; A_e = effective number of haplotypes; A_R = haplotypic richness; H_e = unbiased haplotype diversity with standard errors in brackets; D^2_{sh} is the average genetic distances among individuals and $D^{2}{}_{sh}{}^{ng}$ is the average genetic distances among individuals with no gaps; Ph = number of private haplotypes.

Region	Pop	N	N_h	A_e	A_R	H_e	D^2_{sh}	$D^{2}{}_{sh}{}^{ng}$	Ph
	PG	30	1	1	0	0.00 (0.00)	0	0	0
North	CH	30	3	1.5	2	0.34 (0.10)	0.32	0.13	0
	B	30	3	1.59	2	0.38 (0.09)	10.11	0.78	0
	G	30	4	2.24	3	0.57 (0.05)	5	0.83	0
	AV	31	5	2.28	3.9	0.58 (0.08)	8.52	1.11	2
	ON	30	4	2.6	3	0.64 (0.06)	14.95	1.78	0
Centre	PS	30	4	2.11	3	0.54 (0.07)	1.87	0.49	0
	SF	30	3	1.95	2	0.50 (0.06)	2.03	0.54	1
	A	30	4	1.64	3	0.40 (0.10)	1.06	0.32	0
	V	30	3	1.61	2	0.39 (0.10)	1.37	0.34	1
	SM	30	3	1.15	2	0.13 (0.08)	3.78	0.44	1
	HP	30	5	1.42	4	0.31 (0.11)	2.55	0.37	0
South	BV	30	3	1.85	2	0.48 (0.09)	0.44	0.44	0
	M	30	1	1	0	0.00 (0.00)	0	0	0
	EC	30	6	2.96	5	0.69 (0.08)	1.11	0.79	1
	Mean	30.07	3.47	1.79	2.46	0.4	3.54	0.56	0.4
	SE	0.07	0.35	0.15	1.35	0.06	4.38	0.46	

Population differentiation and Bayesian cluster analysis

The Bayesian clustering suggests the presence of four major groups of populations: a northern cluster, a central cluster with populations from the coast and inland, a southern cluster, and, finally, one cluster with the outlier population SM (Fig 1B). The hierarchical AMOVA showed a stronger differentiation among BAPS clusters than among populations within clusters (Table 4). Contrary to other clusters, in the southern cluster the haplotype H11 is present (Figs 1 and 3). In the central cluster, the haplotype H6 is shared among all populations. The northern cluster had the lowest haplotypic diversity, with only 4 haplotypes and the PG population was fixed for H10 (Table 3).

Finally, the Mantel tests showed an absence of correlation between genetic and geographic distances with both mutation models and thus a lack of isolation by distance pattern. Considering all the populations, the $G_{ST} = 0.29 \pm 0.11$ was among population differentiation based on unordered alleles and the $N_{ST} = 0.28 \pm 0.11$ was based on ordered alleles. The observed N_{ST} was not significantly different from the observed G_{ST}, showing absence of a phylogeographic structure.

Discussion

Spatial distribution of genetic diversity

There is evidence about two possible refugial areas for *A. unedo* in Portugal during the LGM, although it is worth noting that continuous pollen records in the region are missing [47] and that *A. unedo* is a low pollen producer often underrepresented in the pollen diagrams [48]. Nevertheless, Mateus & Queiroz [49] found palaeobotanical evidence of scrublands dominated by *Quercus coccifera* L., including *A. unedo* and other thermophiles shrubs, during the Middle Holocene in the southern region. Santiso *et al.* [5] also postulated a glacial refugium for the species in North Africa-Atlantic Iberia during the LGM. In addition, fossil charcoal records from central Portugal (Estremadura region) support the presence of Mediterranean taxa,
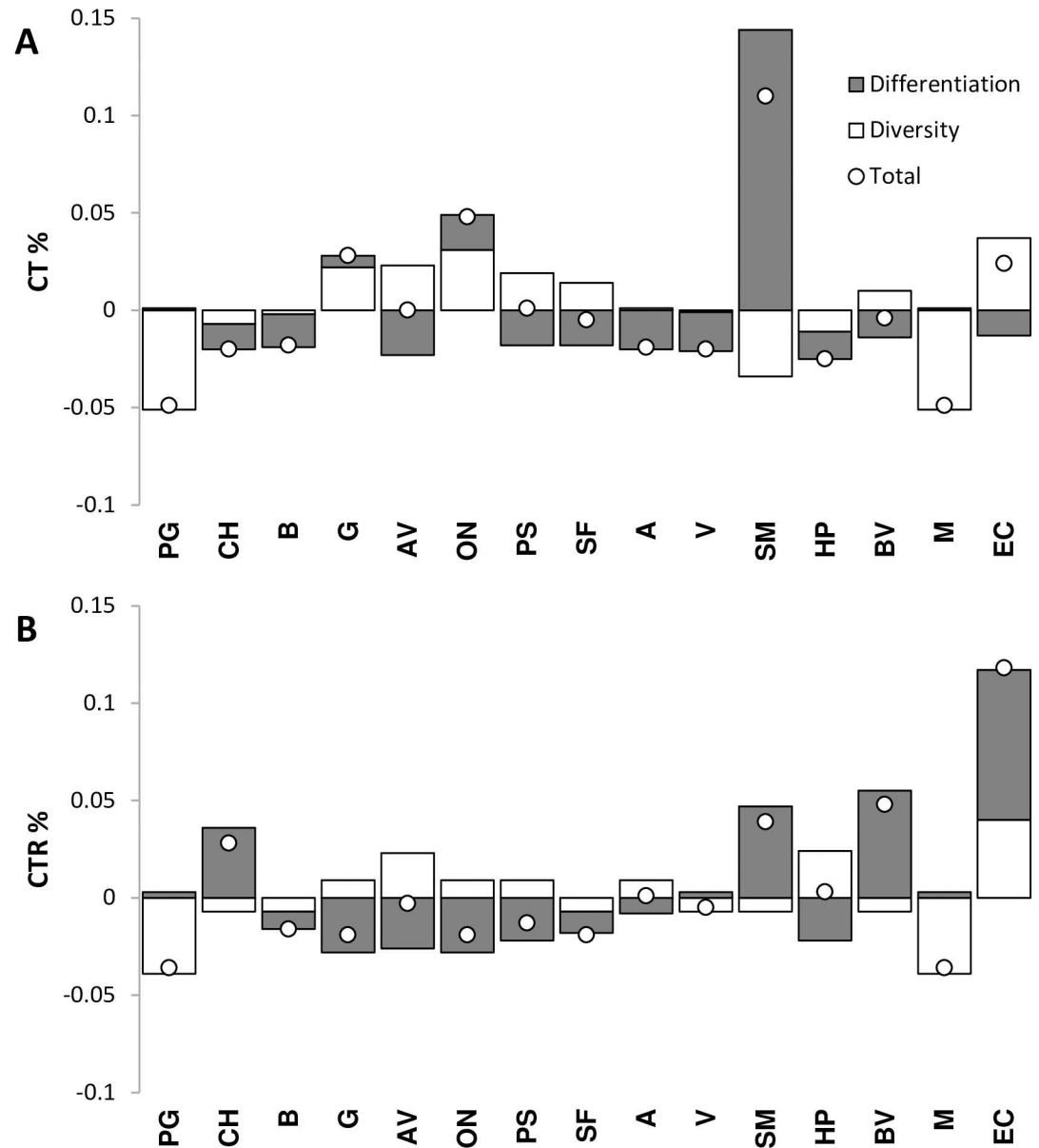
Fig 3. Population diversity partition according to Petit et al. [38] (A) Contribution to the total diversity (CT%) of each population subdivided into own-diversity (white) and divergence (grey) components. (B) Contribution to the total haplotypic richness (CTR%) of each population subdivided into own-diversity (grey) and divergence (white) components.

including *A. unedo*, during the last glacial period [50]. According to these authors, survival of the thermophilous species in central Portugal during the LGM was possible due to the Atlantic Ocean proximity, enough protection from northerly winds and good solar exposure, and average temperatures and precipitations comparable to today. Therefore, it seems likely that most of these species expanded from central Portugal during climate amelioration and that a rapid reforestation occurred after the LGM [51]. In the application of the "refugia within refugia" model to the Iberian Peninsula, the paradigm of Iberia as a single refugium during Pleistocene glacial maxima was challenged, supporting instead the presence of several Iberian refugia for a

Table 4. Analysis of molecular variance (AMOVA) of the *A. unedo* populations. Among all populations (a,c) and among the groups defined by the BAPS analysis (b,d) (hierarchical AMOVA). Distances between haplotypes were measured as number of different alleles (IAM) (a,b) and as number of repeat units for each microsatellite for the i-th locus (SMM) (c,d). SS = sum of squared deviations, d.f. = degrees of freedom, and P = probability of obtaining a more extreme value by chance alone.

Source of Variation	d.f.	SS	Variance Components	% of total variance	Φ statistics[a]	P
(a) All populations						
Among pops	14	39.60	0.09	27.97	$\Phi st = 0.28$	<0.0001
Within pops	436	97.30	0.22	72.03		
Total	450	136.90	0.31			
(b) Hierarchical AMOVA						
Among BAPS clusters	3	29.89	0.09	27.01	$\Phi ct = 0.27$	<0.001
Among pops within clusters	11	9.71	0.02	6.53	$\Phi sc = 0.09$	<0.0001
Within pops	436	97.30	0.22	66.46	$\Phi st = 0.33$	<0.0001
Total	450	136.90	0.33			
(c) All populations						
Among pops	14	353.609	0.81	47.14	$\Phi st = 0.47$	<0.0001
Within pops	436	396.005	0.91	52.86		
Total	450	749.614	1.72			
(d) Hierarchical AMOVA						
Among BAPS clusters	3	310.689	0.99534	49.69	$\Phi ct = 0.50$	<0.005
Among pops within clusters	11	42.920	0.09953	4.97	$\Phi sc = 0.10$	<0.0001
Within pops	436	396.005	0.90827	45.34	$\Phi st = 0.55$	<0.0001
Total	450	749.614				

[a] Fixation indices computed in the Arlequin software [44].

range of flora and fauna [4]. Thus, combining already available knowledge [12] with our results, which are based on a more intensive genetic sampling of *A. unedo* in this area, the hypothesis of the presence of different refugia cannot be excluded. Two well-differentiated genetic clusters were found, one in central and another in southern Portugal. In both clusters, there are populations displaying high genetic diversity (e.g. EC and BV in the southern cluster; G, AV, ON, PS and SF in the central cluster) and some private haplotypes (e.g. high-frequency haplotype H6 is present only in the central and northern cluster). Such findings contrast with the hypothesis of a common postglacial origin of *A. unedo* populations in Portugal.

There are several points about the recent evolutionary history of *A. unedo* at its western-most distribution limit that need to be clarified, and for which the use of genetic markers characterized by higher mutation rates (e.g. nuclear microsatellites) would be helpful. We observed an absence of phylogeographic structure and IBD patterns in our study that could be explained by not having considered the genetic structure found in the analysis. Nevertheless, when tests were repeated only on the central cluster, the only one with sufficient data to obtain meaningful results, the lack of phylogeographic structure and IBD was confirmed (results not shown). Overall, this points towards low levels of gene flow among populations probably due to limitations in the seed dispersal by birds. Whether limited gene flow may increase the effect of genetic drift in erasing IBD patterns, depends on overall gene exchanges among populations and, therefore, on both pollen and seed gene flow [52]. This could be disentangled by assessing both the genetic connectivity through pollen at different spatial scales and by the seed-to-pollen gene flow ratio in *A. unedo* to understand the impact of genetic drift on the spatial distribution of genetic diversity found in the present study.

Another important aspect to be addressed is the effect of contemporary factors on the distribution of genetic variation, with suitable markers. The occurrence of wildfires in the region potentially drives to abrupt changes in population size. Wildfires help in keeping the canopy open, a key factor for *A. unedo* survival and reproduction that prefers open-canopy environments. However if fires are too frequent or intense, the impact on population genetic diversity can be negative, as it is probably the case for the PG population (suffering on average wildfires with 5.7 years interval, since 1975) [53]. We should take into consideration that between 1975–2005, the total area burnt in Portugal was 3,353,000 ha, equivalent to 38% of the country, with a mean annual percent of burn area of 1.2% [22], yet with large inter-annual variability (see S1 Fig). Indeed, increase in fire frequency is seen as a major threat, since short inter-fire periods may exhaust soil-stored and fire-stimulated seed banks, reducing both the effective population size and genetic diversity [54] and eventually increasing the probability of extinction of forest tree populations.

Design of conservation and provenance regions

In the studied region *A. unedo* is, currently, undergoing an intense domestication process, with the commercialisation of improved genetic material (clones) to produce new plantations and the introduction of genetic material through seed from unknown origin [11,17]. Domestication together with growing climate change will affect the species' natural resources. Moreover, the use of vegetative propagation could lead to genetic erosion and inbreeding problems [55], this last effect might also affect the production of fruits, which is the interesting economic product.

Populations are not similar in their capacity to adapt to varying environmental conditions and genetic information can ensure a better use of the available resources by maximizing the evolutionary contribution of the set of populations to conserve [38]. Diversity alone, regardless of the way it is measured, is not a sufficient criterion to identify populations that deserve higher priority for conservation, and when different populations are chosen for conservation, redundancy should be avoided. For conservation purposes the uniqueness of a population, in terms of its allelic composition, may be at least as important as its diversity level, as advised by Petit *et al.* [38]. To select *A. unedo* candidate populations for conservation, besides the genetic diversity parameters, we used the contribution of each population to the total diversity partitioned into two components. The first was related to the level of diversity of the population and the second to its 'uniqueness', *i.e.*, its divergence from the other populations [38].

We found that the SM population, representing a single-population genetic cluster, was the most differentiated, with a high contribution to divergence, but not to diversity levels. We suggest including this population in a conservation programme for *A. unedo* in Portugal since rare haplotypes will be conserved from a putative remnant population. The southern cluster EC population should be also considered as a potential gene reserve, due to its high genetic diversity and haplotypic richness, and for harbouring a private haplotype as SM. This population has the highest contribution to the total haplotypic richness in both the differentiation and diversity components. Similarly, the ON and AV populations, from the central cluster, should also be considered in conservation programs. The first population show the second highest values of genetic diversity and number of haplotypes, besides the average genetic distance among individuals, and the later has the highest haplotypic richness and number of private haplotypes in this cluster. Conservation should include the i) possible impact of very frequent wildfires in the genetic diversity, and the ii) impact of the species domestication at the region scale. In this region, maritime pine was a paradigmatic example of domestication impact in the species genetic structure: the use of seeds from unknown origin has completely

blurred it and affected the species' forest resources [56], and this should be avoided in the strawberry tree case.

Considering the species' genetic improvement, the breeding population used in improvement program, including the clones used for deployment, should mirror the diversity and the haplotypic richness of the species in the region, and care should be taken to monitor the breeding generations ahead. This could be done by comparing the haplotypic and genetic diversity of the breeding population with the current study populations' haplotypic and genetic diversity and by using additional markers, such as nuclear microsatellites and single nucleotide polymorphisms [57,58].

The information retrieved in the current study will also be helpful to design provenance regions aiming at collecting reproductive material for plantations. Those provenance regions should be built based on: i) genetic clustering results, and ii) ecological attributes, including wildfire information [59]. In order to further capture adaptive information, there would be a need for more genetic information, principally about adaptive traits and the genes underlying these traits important both for species preservation and genetic improvement. Direct information about adaptive variation in controlled environments could be done by installing genetic tests, including populations' representative of the clusters detected in the current study. Provenance tests provide information about geographic patters of adaptive genetic variation, which can further help in defining the gene conservation populations. Additionally, they can direct the delineation of breeding zones and seed transfer guidelines, assuring that well adapted genotypes are deployed in artificial regeneration programs [59,60].

Supporting information

S1 Fig. Development of the percentage of burned area (sum of the burned areas/area of the country) and number of wildfires, per year, in the Southern Mediterranean countries, Portugal, Spain, France, Italy and Greece over a 30-year period (1980–2005). (A) Percentage of burned area development. (B) Number of wildfires development. European Commission (2007) Forest Fires in Europe 2006. European Commission, Joint Research Centre. Institute for Environment and Sustainability, Report 7. Luxembourg.
(TIF)

S2 Fig. Sampled populations. Map highlighting the populations' sampled (black dots) and the *A. unedo* stand distribution according to the 2006 Portuguese National Forest Inventory: http://www.icnf.pt/portal/florestas/ifn (orange triangles).
(TIF)

S1 Table. The absolute (counts) and relative (frequency) frequencies of the fifteen detected haplotypes. Haplotype refers to four chloroplast microsatellites composition respectively: AU1, AU2, AU4 and AU7.
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: MMR.

Formal analysis: MMR AP.

Funding acquisition: MMR RC GGV.

Investigation: MMR DG AR.

Project administration: MMR RC GGV.

Resources: MMR RC GGV.

Visualization: MMR AP.

Writing – original draft: MMR AP.

Writing – review & editing: MMR AP LP RC GGV.

References

1. Booy G, Hendriks RJJ, Smulders MJM, Van Groenendael JM, Vosman B (2000) Genetic diversity and the survival of populations. Plant Biol 2: 379–395.

2. Vranckx GUY, Jacquemyn H, Muys B, Honnay O (2012) Meta-analysis of susceptibility of woody plants to loss of genetic diversity through habitat fragmentation. Conserv Biol 26: 228–237. https://doi.org/10.1111/j.1523-1739.2011.01778.x PMID: 22044646

3. Nieto Feliner G (2014) Patterns and processes in plant phylogeography in the Mediterranean Basin. A review. Perspect Plant Ecol Evol Syst 16: 265–278.

4. Gómez A, Lunt DH (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: Weiss S, Ferrand N, editors. Phylogeography of Southern European Refugia: Evolutionary perspectives on the origins and conservation of European biodiversity. Dordrecht: Springer Netherlands. pp. 155–188.

5. Santiso X, Lopez L, Retuerto R, Barreiro R (2016) Phylogeography of a widespread species: Pre-glacial vicariance, refugia, occasional blocking straits and long-distance migrations. AoB Plants 8: plw003, first published online January 1013.

6. Schröter D, Cramer W, Leemans R, Prentice IC, Araújo MB, Arnell NW, et al. (2005) Ecosystem service supply and vulnerability to Gobal Change in Europe. Science 310: 1333–1337. https://doi.org/10.1126/science.1115233 PMID: 16254151

7. Torres JA, Valle F, Pinto C, García-Fuentes A, Salazar C, Cano E (2002) *Arbutus unedo* L. communities in southern Iberian Peninsula mountains. Plant Ecol 160: 207–223.

8. Takrouni MM, Ali IBEH, Messaoued C, Boussaid M (2012) Genetic variability of Tunisian wild strawberry tree (*Arbutus unedo* L.) populations interfered from isozyme markers. Sci Hortic 146: 92–98.

9. Takrouni MM, Boussaid M (2010) Genetic diversity and population's structure in Tunisian strawberry tree (*Arbutus unedo* L.). Sci Hortic 126: 330–337.

10. Lopes L, Sá O, Pereira JA, Baptista P (2012) Genetic diversity of Portuguese *Arbutus unedo* L. populations using leaf traits and molecular markers: An approach for conservation purposes. Sci Hortic 142: 57–67.

11. Gomes F, Costa R, Ribeiro MM, Figueiredo E, Canhoto JM (2013) Analysis of genetic relationship among *Arbutus unedo* L. genotypes using RAPD and SSR markers. J For Res 24: 227–236.

12. Santiso X, Lopez L, Retuerto R, Barreiro Lozano R (2016) Population structure of a widespread species under balancing selection: the case of *Arbutus unedo* L. Front Plant Sci 6.

13. Celikel G, Demirsoy L, Demirsoy H (2008) The strawberry tree (*Arbutus unedo* L.) selection in Turkey. Sci Hortic 118: 115–119.

14. Pedro JG (1994) Portugal Atlas do Ambiente. Lisboa: D.G.A., Ministério do Ambiente e Recursos Naturais.

15. Gomes F, Suárez D, Santos R, Silva M, Gaspar D, Machado H (2016) Mycorrhizal synthesis between *Lactarius deliciosus* and *Arbutus unedo* L. Mycorrhiza 26: 177–188. https://doi.org/10.1007/s00572-015-0656-1 PMID: 26215114

16. Miguel MG, Faleiro ML, Guerreiro AC, Antunes MD (2014) *Arbutus unedo* L.: chemical and biological properties. Molecules 19: 15799–15823. https://doi.org/10.3390/molecules191015799 PMID: 25271425

17. Gomes F, Simões M, Lopes ML, Canhoto JM (2010) Effect of plant growth regulators and genotype on the micropropagation of adult trees of *Arbutus unedo* L. (strawberry tree). N Biotechnol 27: 882–892. https://doi.org/10.1016/j.nbt.2010.02.009 PMID: 20219713

18. FAO (2010) The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture. Rome: Food and Agriculture Organization.

19. Lefèvre F, Boivin T, Bontemps A, Courbet F, Davi H, Durand-Gillmann M, et al. (2014) Considering evolutionary processes in adaptive forestry. Ann For Sci 71: 723–739.

20. Fady B (2012) Biogeography of neutral genes and recent evolutionary history of pines in the Mediterranean Basin. Ann For Sci 69: 421–428.

21. Dury M, Hambuckers A, Warnant P, Henrot A, Favre E, Ouberdous M, et al. (2011) Responses of European forest ecosystems to 21(st) century climate: assessing changes in interannual variability and fire intensity. iForest 4: 82–99.

22. Oliveira SLJ, Pereira JMC, Carreiras JMB (2012) Fire frequency analysis in Portugal (1975–2005), using Landsat-based burnt area maps. Int J Wildland Fire 21: 48–60.

23. Pausas JG, Keeley JE (2009) A burning story: The role of fire in the history of life. BioScience 59: 593–601.

24. Rodríguez-Sánchez F, Pérez-Barrales R, Ojeda F, Vargas P, Arroyo J (2008) The Strait of Gibraltar as a melting pot for plant biodiversity. Quat Sci Rev 27: 2100–2117.

25. Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Mol Ecol 14: 689–701. https://doi.org/10.1111/j.1365-294X.2004.02410.x PMID: 15723661

26. Sealy JR (1949) Arbutus unedo. J Ecol 37: 365–388.

27. Herrera CM (1998) Long-term dynamics of Mediterranean frugivorous birds and fleshy fruits: a 12-year study. Ecol Monogr 68: 511–538.

28. Rasmont P, Regali A, Ings TC, Lognay G, Baudart E, Marlier M, et al. (2005) Analysis of pollen and nectar of *Arbutus unedo* as a food source for *Bombus terrestris* (Hymenoptera: Apidae). J Econ Entomol 98: 656–663. PMID: 16022289

29. Mesléard F, Lepart J (1989) Continuous basal sprouting from a lignotuber: *Arbutus unedo* L. and *Erica arborea* L., as woody Mediterranean examples. Oecologia 80: 127–131. https://doi.org/10.1007/BF00789941 PMID: 23494355

30. Panagos P, Van Liedekerke M, Jones A, Montanarella L (2012) European Soil Data Centre: Response to European policy support and public data requirements. Land Use Policy 29: 329–338.

31. Couto MAG, editor (2011) Iberian climate atlas: air temperature and precipitation (1971–2000). Madrid: Agencia Estatal de Meteorologia and Instituto de Meteorologia de Portugal.

32. Martinez-Alberola F, del Campo EM, Lazaro-Gimeno D, Mezquita-Claramonte S, Molins A, Mateu-Andres I, et al. (2013) Balanced gene losses, duplications and intensive rearrangements led to an unusual regularly sized genome in *Arbutus unedo* chloroplasts. PLoS One 8: e79685. https://doi.org/10.1371/journal.pone.0079685 PMID: 24260278

33. Kofler R, Schlötterer C, Lelley T (2007) SciRoKo: a new tool for whole genome microsatellite search and investigation. Bioinformatics 23: 1683–1685. https://doi.org/10.1093/bioinformatics/btm157 PMID: 17463017

34. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. (2012) Primer3—new capabilities and interfaces. Nucl Acids Res 40: e115. https://doi.org/10.1093/nar/gks596 PMID: 22730293

35. Paradis E (2010) pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics 26: 419–420. https://doi.org/10.1093/bioinformatics/btp696 PMID: 20080509

36. R-Core-Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna. http://www.R-project.org/.

37. Nei M (1987) Molecular evolutionary genetics. New York: Columbia University Press. 333 p.

38. Petit R, El-Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. Conserv Biol 12: 844–855.

39. Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28: 2537–2539. https://doi.org/10.1093/bioinformatics/bts460 PMID: 22820204

40. Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139: 463–471. PMID: 7705647

41. Vendramin GG, Anzidei M, Madaghiele A, Bucci G (1998) Distribution of genetic diversity in *Pinus pinaster* Ait. as revealed by chloroplast microsatellites. Theor Appl Genet 97: 456–463.

42. Corander J, Marttinen P, Siren J, Tang J (2008) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. BMC Bioinformatics 9: 539. https://doi.org/10.1186/1471-2105-9-539 PMID: 19087322

43. Bonin A, Ehrich D, Manel S (2007) Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. Mol Ecol 16: 3737–3758. https://doi.org/10.1111/j.1365-294X.2007.03435.x PMID: 17850542

44. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Eco Resour 10: 564–567.

45. Heuertz M, Teufel J, González-Martínez SC, Soto A, Fady B, Alía R, et al. (2010) Geography determines genetic relationships between species of mountain pine (Pinus mugo complex) in western Europe. J Biogeogr 37: 541–556.

46. Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. Genetics 144: 1237–1245. PMID: 8913764

47. Fletcher WJ, Boski T, Moura D (2007) Palynological evidence for environmental and climatic change in the lower Guadiana valley, Portugal, during the last 13 000 years. The Holocene 17: 481–494.

48. Mateus JE (1992) Holocene and present-day ecosystems of the Carvalhal Region, Southwest Portugal. PhD thesis. Utreque: Universidade de Utreque. 184 p.

49. Queiroz PF, Mateus JE (2011) Estudo polínico dos tufos calcários da Ribeira da Asseca, Tavira. Relatório Final Projecto HOLOCLIMA PTDC/CTE-GEX/71298/2006. Lisbon.

50. Figueiral I, Terral JF (2002) Late Quaternary refugia of Mediterranean taxa in the Portuguese Estremadura: charcoal based palaeovegetation and climatic reconstruction. Quat Sci Rev 21: 549–558.

51. Santos L, Goñi MFS (2003) Lateglacial and Holocene environmental changes in Portuguese coastal lagoons 3: vegetation history of the Santo Andre coastal area. The Holocene 13: 459–464.

52. Ouborg NJ, Vergeer P, Mix C (2006) The rough edges of the conservation genetics paradigm for plants. J Ecol 94: 1233–1248.

53. Vincenzi S, Piotti A (2014) Evolution of serotiny in maritime pine (Pinus pinaster) in the light of increasing frequency of fires. Plant Ecol 215: 689–701.

54. Pausas JG, Keeley JE (2014) Evolutionary ecology of resprouting and seeding in fire-prone ecosystems. New Phytol 204: 55–65. PMID: 25298997

55. Lefèvre F (2004) Human impacts on forest genetic resources in the temperate zone: an updated review. For Ecol Manag 197: 257–271.

56. Ribeiro MM, Plomion C, Petit R, Vendramin GG, Szmidt AE (2001) Variation in chloroplast single-sequence repeats in Portuguese maritime pine (Pinus pinaster Ait.). Theor Appl Genet 102: 97–103.

57. El-Kassaby YA (1995) Evaluation of the tree-improvement delivery system: factors affecting genetic potential. Tree Physiol 15: 545–550. PMID: 14965941

58. Freeman JS, Marques CMP, Carocha V, Borralho N, Potts BM, Vaillancourt RE (2007) Origins and diversity of the Portuguese Landrace of Eucalyptus globulus. Ann For Sci 64: 639–647.

59. White TL, Adams WT, Neale DB (2007) Forest Genetics Oxfordshire, UK: CAB International. 704 p.

60. Lopez GA, Potts BM, Dutkowski GW, Rodriguez Traverso JM (2001) Quantitative genetics of Eucalyptus globulus: affinities of landrace and native stand localities. Silvae Genet 50: 5–6.