Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA)
Available online at www.inia.es/forestsystems

Forest Systems 2010 19(2), 156-169
ISSN: 2171-5068
eISSN: 2171-9845

Genetic structure of wild Spanish populations of *Castanea sativa* as revealed by isozyme analysis

J. Fernández-López[1]* and A. B. Monteagudo[1,2]

[1] *Centro de Investigación Forestal de Lourizán. Consellería do Medio Rural. Xunta de Galicia. P.O. Box 127. 36080 Pontevedra. Spain*

[2] *Actual address: Centro de Investigaciones Agrarias de Mabegondo. INGACAL. Consellería de Medio Rural. Xunta de Galicia. Spain*

Abstract

The genetic variability within and among 17 wild Spanish chestnut stands was examined by isozyme analysis, with the goals of describing their geographic structure and designing conservation and management strategies. Measures of genetic diversity such as allelic richness, heterozygosity, polymorphism, F-statistics, D-statistics, gene flow and the contributions of each stand to diversity and allelic richness were calculated, clustering using Nei's genetic distances and a method based on an Monte Carlo Markov Chain algorithm were used.

Wild Spanish chestnut populations displayed heterogeneity of allele frequencies between them, high levels of genetic diversity and high differentiation ($Fst = 0.15$) compared with populations from other western European countries. Two clustering methods allowed identification of three clusters: The highest heterozygosity and allelic richness were found in the North and especially in the Galician cluster close to Portugal. These results indicate that probably the northern chestnut populations are relictual originated from a North Iberian refuge. Several areas can be recommended for inclusion in the network of Conservation Units: Fraga de Catasós, representing the southern Galician cluster; Fragas do Eume, representing the northern Spanish cluster; and Hervás or El Tiemblo, representing the Mediterranean cluster.

Key words: relict populations, diversity, allelic richness, conservation.

Resumen

Determinación de la estructura genética de las poblaciones silvestres españolas de *Castanea sativa* mediante isoenzimas

La variabilidad genética entre y dentro de 17 rodales de castaño silvestre muestreados en poblaciones españolas se examinó mediante el análisis de polimorfismos isoenzimáticos con la finalidad de describir la estructura geográfica y diseñar estrategias de conservación y manejo. Se calcularon la riqueza alélica, heterocigosidad, polimorfismo, la estadística F y D, el flujo genético y se realizaron análisis de agrupación de poblaciones utilizando la distancia genética de Nei y un método basado en el algoritmo Monte Carlo Markov Chain, y finalmente se calcularon las contribuciones de cada rodal a la diversidad y riqueza alélica.

Las poblaciones silvestres de castaño españolas presentaron heterogeneidad de frecuencias alélicas entre ellas, elevados niveles de diversidad genética y de diferenciación ($Fst = 0{,}15$) comparadas con otras poblaciones del Oeste de Europa. Se identificaron tres grupos de poblaciones: el grupo Mediterráneo, el grupo del Norte y el grupo del Sur de Galicia. La mayor heterocigosidad y riqueza alélica está en el Norte, principalmente en el grupo del Sur de Galicia. Estos resultados indican que las actuales poblaciones de castaño del Norte de la península Ibérica son poblaciones relícticas originadas en uno o varios refugios situados en esta área. Se recomiendan varias áreas para incluir en la red de Unidades de Conservación: Fraga de Catasós, en representación del grupo Sur de Galicia; las Fragas do Eume, representando el grupo del Norte; y Hervás o el Tiemblo en representación del grupo Mediterráneo.

Palabras clave: poblaciones relícticas, diversidad, riqueza alélica, conservación.

* Corresponding author: josefa.fernandez.lopez@xunta.es
Received: 03-04-09; Accepted: 06-03-10.

Introduction

The widespread and ancient management system of the sweet chestnut (*Castanea sativa* Miller) across its distribution area led geneticists and experts in forest management to consider that the current genetic structure of Iberian populations is more affected by the introduction of germplasm from other areas and by artificial selection than by natural evolution. In fact, most of the literature regarding Iberian chestnut resources considers them to be naturalised or landraces and not natural and autochthonous. Although human influence due to management is clear, there is evidence from fossil remains of the conservation of some relictic populations during glaciation periods (Krebs *et al.,* 2004). These relictic populations may be the source of the current chestnut populations.

The sweet chestnut was long considered to have been introduced into the Iberian Peninsula by the Romans to be cultivated for nut and wood production. Historical evidence of chestnut cultivation during the Roman period, however, is limited to northern Italy and indicates that the development of chestnut culture in Western Europe for nut production occurred during medieval times (Conedera *et al*., 2004).

The presence of pollen and other fossil remains have revealed the pre-Roman existence of *Castanea* in northern regions of the Iberian Peninsula (Martín-Arroyo *et al.,* 1999; Carrión *et al*., 2003), as was previously suggested by Zohary and Hopf (1988) and by Huntley and Birks (1983). Krebs *et al*. (2004) found palynological evidence to support the presence of two refugia during the last glaciation in the Iberian Peninsula, one in the Cantabrian coast and another in the northwestern area between southern Galicia (Spain) and northern Portugal. In a recent study involving predictive modelling of the distribution of forest tree species, Benito Garzón *et al*. (2007) suggested the foothills of the Pyrenees and the northwest of the Iberian Peninsula as the most likely refugia for species during the last glaciation

Studies involving isozyme analysis have identified eastern Turkish sweet chestnut demes as the founder stock of European chestnut populations following glaciation because of their high genetic diversity, which decreases from eastern Turkey to Italy and France (Villani *et al*., 1994), with the levels of genetic diversity being higher in French than in Italian populations (Bonnefoi, 1984; Villani *et al*., 1991a; Machon *et al*., 1996). The Greek gene pool, identified as the most diverse in Europe, has probably contributed the most to western populations (Aravanopoulos *et al.,* 2001, 2005). These studies, however, did not include Iberian populations.

The current distribution of the chestnut in Spain is discontinuous, with the largest surfaces located between Galicia and Navarra, along with several discontinuous patches in the central western Spanish and Andalucian mountains (Fig. 1). This species usually grows together with other species, although it also grows in monospecific stands, usually managed as coppices or as orchards. The massive mortality of chestnuts in southwestern Europe caused by different species of *Phytophthora* led to the introduction of seeds of Japanese (*Castanea crenata* Sieb. et Zucc.) and Chinese chestnuts (*Castanea mollissima* Blume). Since 1940, two hybridisation programmes were developed in Galicia to breed sweet chestnuts for resistance to ink disease caused by *Phytophthora* spp. Most of the hybrids were crosses between the sweet chestnut and Japanese chestnut; the majority were F_1 and open-pollinated progeny of F_1 and F_2 and were used in plantations in northern Spain over the last five decades.

In this study, isoenzymatic genetic markers were used to assess the genetic variability of wild Spa-

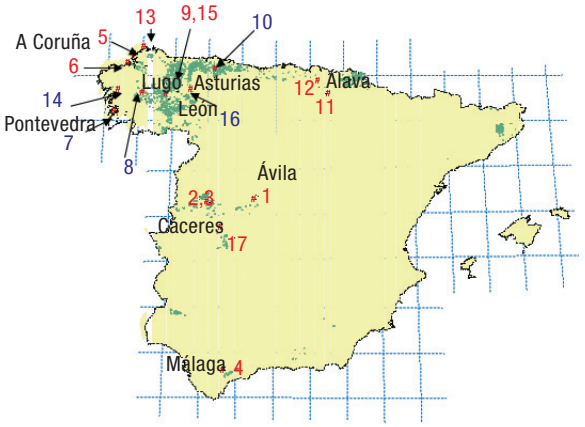

Figure 1. Geographic distribution of the studied populations. The green area represents the current distribution of sweet chestnut stands in Spain. The numbers of sampled stands are coloured according to their clustering into three ancestral populations identified with STRUCTURE. Stands with numbers in red belong to cluster $Cl1_{STR}$, comprising the Mediterranean populations and two stands in the eastern Lugo mountains. Stands shown in green numbers belong to cluster $Cl2_{STR}$, comprising the northern Galician and Basque Country populations. Stands shown in blue numbers belong to cluster $Cl3_{STR}$, comprising populations in Pontevedra, the Lugo plateau and northeastern León. Stands 10 and 17 were identified as hybrids between the different identified clusters.

nish chestnut populations in order to describe their genetic structure and understand more about their evolution in southern Europe, with the goals of designing adapted strategies for conservation and management.

Materials and methods

Populations studied

A total of 512 trees collected in 17 wild stands representing most of the Spanish distribution area of *C. sativa* were studied. Twelve sites were selected in the northern region, the largest continuous extension of the distribution area; four sites were selected in the central region; and another one was selected in the extreme southern area of the distribution (Table 1, Fig. 1). The sampling was designed on the basis of the results of previous adaptive traits studies of chestnut populations (Fernández-López *et al*., 2005). In stands 8, 9 and 15, there was evidence of a possible effect of domestication due to the presence of grafts, which were avoided. Sampling was carried out in stands in which there were no hybrids between Japanese chestnut and sweet chestnut. The form and width of the last growth shoot section as well as bud burst date allow the identification of hybrids.

Isozyme analysis

Ten isozyme systems encoded by 13 loci were analysed: alcohol dehydrogenase (*Adh*, E.C.1.1.1.1), phosphoglucomutase (*Pgm-2*, E.C.2.7.5.1), glucose phosphate isomerase (*Pgi-2*, E.C.5.3.1.9), malate dehydrogenase (*Mdh-2*, Mdh-3, E.C.1.1.1.37), peroxidase (*Prx-2*, E.C.1.11.1.7), 6-phosphoglucose dehydrogenase (*6Pgd-2*, E.C.1.1.1.44), diaphorase (*Dia-1*, *Dia-2*, E.C.1.6.4.3), UTP-glucose-1-phosphate-uridyltransferase (*Ugp-1*, E.C.2.7.7.9), shikimate dehydrogenase (*Sdh*, E.C.1.1.1.25) and isocitric dehydrogenase (*Idh-1*, *Idh-2*, E.C.1.1.1.42). These loci have been used previously for chestnut species (Malvolti and Fineschi, 1987; Villani *et al*., 1991a; Huang *et al*., 1994; Fernández-López, 1996). The isozyme loci *Pgi-2*, *Pgm-2*, *Sdh*, *Mdh-2, Mdh-3* and *Pgd-2* were used to verify that germplasm of the Japanese or Chinese chestnut was not present in the samples. Among these loci only *Pgm-2* and *Sdh* show diagnostic loci, while *Pgi-2*, *Mdh-2*, *Mdh-3* and *Pgd-2* show species specific alleles not fixed, although the allele 105 of *Pgi-2* is almost fixed in Japanesse chestnut (Fernández-López, 1996).

Samples from 27-37 trees per stand were analysed by starch gel electrophoresis (Table 1). Two additional samples (M_1, M_2) were introduced in each gel to help identify the bands and to exclude the presence of Asiatic germplasm. M_1 was *Castanea sativa*, and M_2 was an

Table 1. The wild chestnut stands sampled and their geographical origin

Region	Stand	Locality	Province	Coordinates		Altitude (m)	N*
				Longitude	Latitude		
Mediterranean	1	El Tiemblo	Ávila	04°31'57"W	40°20'43"N	1,300	30
	2	Hervás	Cáceres	05°51'52"W	40°15'13"N	980	30
	3	Hervás	Cáceres	05°52'31"W	40°15'21"N	880	30
	4	Ronda	Málaga	05°18'04"W	36°32'14"N	720	37
	17	Guadalupe	Cáceres	05°20'00"W	39°28'00"N	950	29
Atlantic Galician	5	Maniños	A Coruña	08°11'41"W	43°26'39"N	100	30
	6	Sigrás	A Coruña	08°21'47"W	43°17'07"N	100	27
	7	San Cibrán	Pontevedra	08°40'56"W	42°11'17"N	270	30
	13	Eume	A Coruña	08°02'16"W	43°24'17"N	214	30
North and Galician plateau	8	Chantada	Lugo	07°51'60"W	42°36'32"N	680	29
	9	Mercurín	Lugo	07°10'16"W	42°37'51"N	830	30
	10	Nandiello	Asturias	05°45'5"W	43°13'04"N	400	30
	11	Antoñana	Álava	02°24'28"W	42°41'48"N	735	30
	12	Etxaguen	Álava	02°43'30"W	42°59'30"N	680	30
	14	Catasós	Pontevedra	08°05'32"W	42°38'12"N	580	30
	15	Mercurín	Lugo	07°09'5"W	42°38'18"N	650	30
	16	Médulas	León	06°46'10"W	42°27'25"N	850	30

*N: number of trees sampled within each stand.

F_1 hybrid between European and Japanese chestnut. Buds were homogenised with cold Tris-HCl extraction buffer, pH 7.3, according to Kim (1979). Whatman no. 3 filter papers (3 × 8 mm) were soaked in protein extracts and loaded into a cut in the 10% starch gels. Different gel buffer systems were used depending on the enzyme loci: histidine ($H_{7.0}$), histidine-citrate ($HC_{5.7}$) (according to Shields *et al.,* 1983) and morpholine-citrate 6.1 ($MC_{6.1}$) (as described by Wendel and Weeden, 1989).

Data analysis

Genetic variation within stands was analysed by allelic frequencies, and the G^2 test for homogeneity of gene frequencies across stands (Nei, 1987) was also carried out. Genetic structure was estimated using F-statistics, with F_{st}, F_{it}, F_{is} (Weir, 1990) and the allelic fixation index, F_{is} (Wright, 1978). The Ewens-Watterson test for neutrality (Manly, 1985) was performed for each locus to detect possible effects of selection on the inter-population allele distribution.

The genetic distances among subpopulations were estimated according to Nei (1978), and a cluster analysis was performed by the UPGMA method. The POPGENE 1.31 (Yeh and Boyle, 1997) and NTSYS 2.1 (Rohlf, 2000) software packages were used for this purpose. The Bayesian method implemented in the software STRUCTURE was used for the identification of a *K* number of ancestral populations from which the current populations originated (Pritchard *et al.,* 2000). A no admixture model with a burning period of 40,000 followed by 40,000 iterations was used for *K* values from 2 to 6. Ten independent runs were carried out and further aligned with the software CLUMPP version 1.1.1 (Jakobsson and Rosenberg, 2007). A histogram displaying the ancestry of each individual in each ancestral population was constructed with the software DISTRUCT (Rosenberg, 2004). Clustering of a subpopulation to an ancestral population was done considering the mean ancestry of the individuals within the respective subpopulation in each $K = 3$ ancestral population. A subpopulation was considered to be included within a cluster if the mean ancestry in the corresponding population was ≥ 0.50, and the subpopulations with their ancestry value in either ancestral population < 0.50 were considered to be hybrids among the different demes.

The two-locus population subdivision (D-statistics) analysis proposed by Ohta (1982a, 1982b) was also applied. In this analysis, the total variance of dilocus linkage disequilibrium (*LD*) is partitioned into D^2_{IS} (the variance of within subpopulation disequilibrium), D^2_{ST} (the variance of the correlation of genes of the two loci of different gametes in a subpopulation relative to that of the total population), D'^2_{IS} (the variance of the correlation of genes of the two loci of one gamete in a subpopulation relative to that of the total population) and D'^2_{ST} (the variance of the disequilibrium of the total population). According to Ohta, *LD* is due to limited migration (drift) if $D^2_{ST} > D^2_{IS}$ and $D'^2_{IS} > D'^2_{ST}$, and epistatic natural selection is a more likely cause if $D^2_{ST} < D^2_{IS}$ and $D'^2_{IS} < D'^2_{ST}$.

The number of observed alleles (n_a), the effective number of alleles (n_e) (Hartl and Clark, 1989), the observed and expected heterozygosity (H_o and H_e) (Levene, 1949; Nei, 1973) and the percentage of polymorphic loci (*P*) were computed on the basis of a 5% criterion for each subpopulation. The heterozygote deficit within each stand was estimated as $F_{is} = 1 - (H_o/H_e)$. The coefficient of relative differentiation in allelic richness for each locus (*Rst*), allelic richness (*Ar*) and the contributions of each subpopulation to the total Nei diversity (*Ct*) and the total allelic richness (*Crt*) were estimated according to the methods described in Petit *et al.* (1998). The allelic richness was standardised to 27, the smallest sample size. The contribution of each subpopulation was divided in two components, the contribution due to its own diversity and allelic richness (*Cs/Crs*) and the contribution due to its own divergence in diversity and allelic richness (*Cd/Crd*).

The value of F_{st} was also computed by excluding each of the subpopulations in turn: a reduction in the value of F_{st} is expected when a differentiated subpopulation is excluded from the calculation (Caccone and Sbordoni, 1987). Gene flow between subpopulations and groups of subpopulations was estimated as $N_m = 0.25(1-F_{st})/F_{st}$ (Slatkin and Barton, 1989). This value was also computed by excluding each subpopulation in turn, as for F_{st}. An increase in the N_m value is expected when a differentiated subpopulation is removed from the computation.

Results

Eleven polymorphic loci were identified (Table A.1). The alleles $Pgi\text{-}2^{105}$, Sdh^{97}, $Idh\text{-}1^{108}$, $Idh\text{-}2^{105}$, $Ugp\text{-}1^{100}$ and $Mdh\text{-}3^{95}$ were absent from some subpopulations or present at a frequency lower than 0.1. Only one tree in

the stand 8 was identified as a hybrid between *C. crenata* and *C. sativa*, and it was removed from further analysis.

Most of the subpopulations in central and southern Spain displayed an absence of low frequency alleles, with the exception of *Idh-2*^{105}, which occurred at the highest frequencies in two stands from central Spain. The alleles *Pgi-2*^{105} and *Sdh*^{97} were mainly present in the stands in northwestern Spain. The results of the Ewens-Watterson test for neutrality (not shown) determined a low probability of this occurring by chance, indicating possible effects of selection on the distribution of the allelic richness of the *Pgi-2* and *Sdh* loci. *Idh-1*^{100} was fixed only in subpopulation 4; *Idh2*^{100} was fixed in the three northern subpopulations 10, 11 and 12; *Ugp-1*^{95} was fixed in subpopulations 1, 4 and 15; *Mdh-3*^{100} was fixed in subpopulations 1 and 4 and most northern subpopulations (6, 13, 11 and 8). On the other hand, subpopulations 1, 4, and 11 displayed the greatest number of fixed alleles, while subpopulations 5, 7, 14 and 16 had no fixed alleles for any of the polymorphic loci. The results of the G test showed heterogeneity of gene frequencies across subpopulations for all loci.

Hardy-Weinberg (H-W*)* equilibrium was followed in the majority of the populations for most of the loci (data not shown). Significant departures from H-W equilibrium were observed primarily due to a deficit in heterozygotes. The values of Wright's fixation index *(F_{is}* Wright) (Table 2) for the *Sdh* and *Idh-1* loci displayed a deviation from zero towards a high and statistically significant deficit of heterozygotes, consistent with values for low gene flow (1.054 and 0.598), high values for the F_{it} index (0.336 and 0.228) and high values for the F_{st} index (0.192 and 0.295). The *Pgi-2* and *Dia-2* loci showed a deviation from zero towards a statistically significant deficit of heterozygotes. In this case, the gene flow was higher (1.240 and 1.409), and the values of F_{is} and F_{st} were lower than the values for *Sdh* and *Idh-1*. The *Prx-2* locus displayed a statistically significant deficit in heterozygotes, with a low value of F_{st} (0.074) and high gene flow (3.149). The *Pgi-2*, *Sdh*, *Idh-1* and *Dia-1* loci had higher values in comparison with the global mean F_{st} value (0.152). These loci, therefore, had contributed more to the differentiation between subpopulations, while the *Idh-2*, *Ugp-1* and *Prx-2* loci made little contribution to the differentiation of the subpopulations.

The differentiation among subpopulations was also reflected in the genetic distances (Table A.2). Subpopulation 11 displayed the greatest identified distances with subpopulations 1, 2, 9, 15 and 16. Subpopulation 14 displayed high genetic distance from subpopulations 1, 4 and 8. Subpopulations from central Spain (1, 2, 3) were genetically closer, as were North-western Galician subpopulations (5/6/13), subpopulations from Eastern Galician mountains (15/9) and subpopulations 17/4, which were geographically closer to each other (Fig. 1). The genetic distances were used to construct a *UPGMA* dendrogram (Fig. 2) in which three fundamental groups were differentiated. Cluster 1 ($Cl1_{UPGMA}$) included a wide geographic zone: northwestern (stands 9, 10, 15 and 16), central (stands 1, 2, 3 and 17) and southern (stand 4) Spain. Cluster 2 ($Cl2_{UPGMA}$) included some subpopulations on the northwestern coast of Spain (stands 6, 5, 13) and two stands in the North (11 and

Table 2. Summary of the Wright fixation index, *F*-statistics and gene flow for wild chestnut populations

Loci	F_{is} wright†		*Fis*	*Fit*	*Fst*	*Nm*	*Rst*
Pgi-2	0.0519	**	–0.143	0.049	0.168	1.240	0.16
Sdh	0.3419	**	0.179	0.336	0.192	1.054	0.12
Idh-1	0.2356	**	–0.096	0.228	0.295	0.598	0.06
Idh-2	0.0215		–0.068	0.016	0.079	2.906	0.13
Ugp-1	–0.0781		–0.175	–0.082	0.079	2.904	0.17
Mdh-2	0.0606		–0.059	0.061	0.113	1.963	0.00
Mdh-3	0.0074		–0.131	0.007	0.122	1.804	0.36
Adh	0.0618		–0.024	0.061	0.083	2.752	0.00
Prx-2	0.1860	**	0.120	0.185	0.074	3.149	0.00
Dia-1	0.0775		–0.152	0.075	0.197	1.018	0.00
Dia-2	0.0910	*	–0.072	0.090	0.151	1.409	0.00
Mean	0.0961		–0.036	0.121	0.152	1.392	0.09

†* $P<0.05$. ** $P<0.01$. *Fis:* inbreeding coefficient at stand level. *Fit:* inbreeding coefficient at the total sample level. *Fst:* proportion of differentiation among stands. *Nm:* gene flow. Rst: relative differentiation in allelic richness.

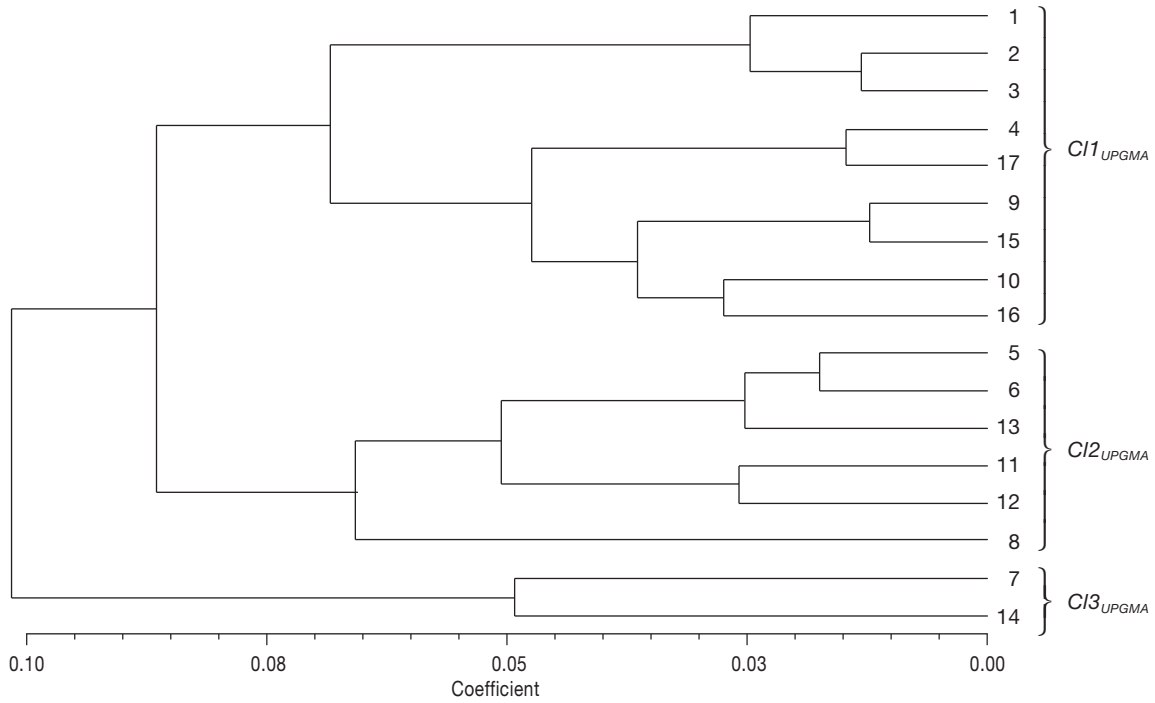

Figure 2. The UPGMA dendrogram of wild chestnut populations based on Nei's unbiased genetic distances (1978) classified the populations into three clusters. Stands within each cluster are spatially closer among themselves than with stands in other clusters.

12). Cluster 3 ($Cl3_{UPGMA}$) comprised the subpopulations on the southwestern coast of Pontevedra (stands 7 and 14), in northwest Spain and close to northern Portugal.

The output of STRUCTURE for ten runs for $K = 3$ showed high repetition, as indicated by the results of CLUMPP, which gave a very high average pairwise similarity value ($H' = 0.99$). In Figure 3, each individual is represented by a vertical bar with the proportion of each ancestral population in each individual, and Table 3 contains the mean ancestry of each subpopulation in each $K = 3$ populations. The classification of subpopulations based on their mean ancestry in each ancestral population was similar to *UPGMA* clustering but with some differences. Cluster 1 ($Cl1_{STR}$) was similar to $Cl1_{UPGMA}$ but lost subpopulations 16 and 10. Subpopulations 9, 15 and 17 in $Cl1_{UPGMA}$ had their main ancestry in $Cl1_{STR}$ but had a certain proportion of their ancestry in $Cl3_{STR}$ (in the case of subpopulations 9 and 15) and a mixture of the three $Cl1\text{-}2\text{-}3_{STR}$ (in the case of subpopulation 17). Subpopulations 10 and 16, included in $Cl1_{UPGMA}$, had their main percentage of ancestry in the southern Galician cluster, although they were hybrid populations between the three ancestral populations and between $Cl3_{STR}$ and $Cl1_{STR}$, respectively. Subpopulation 8, which was in $Cl1_{UPGMA}$, is a hybrid population between $Cl2_{STR}$ and $Cl3_{STR}$. The main difference between the two northern ancestral populations $Cl2_{STR}$ and $Cl3_{STR}$ was in the alleles $Pgi\text{-}2^{105}$, $Mdh\text{-}3^{95}$ and $Dia2^{100}$, which were almost absent in the cluster $Cl2_{STR}$. The main difference between the $Cl1_{STR}$ and the

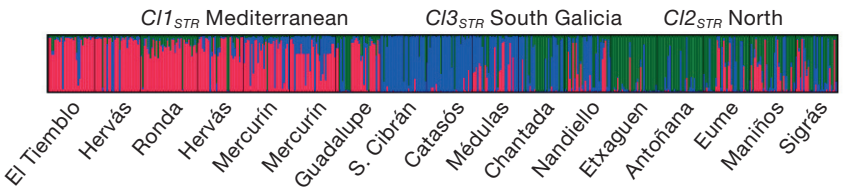

Figure 3. Bayesian clustering using STRUCTURE of 513 individuals collected in 17 stands genotyped with 11 isozyme loci. Each individual is represented by a vertical bar, and the different colours represent membership coefficients in the three different clusters: the red colour represent the cluster dominant in central and southern Spain in the Mediterranean deme ($Cl1_{STR}$); the blue colour represents the cluster dominant in the southern and central Galician populations and in León ($Cl3_{STR}$); and the green colour represents the cluster dominant in the northern and northwestern populations close to the sea ($Cl2_{STR}$).

Table 3. Mean ancestry of each stand in each identified population: $Cl1_{STR}$, $Cl2_{STR}$, and $Cl3_{STR}$. Stands were assigned to one of the three groups —Mediterranean, southern Galician and northern Spanish— depending on the population in which the stand had its main ancestry. Within each group, the populations were grouped by decreasing ancestry value in the main respective population. Populations 10 and 17 were considered hybrid between different populations. The sum of ancestries of each stand was 1

Cluster region	Stand	Ancestry		
		Cl1	*Cl2*	*Cl3*
$Cl1_{STR}$	1	0.82	0.10	0.08
Mediterranean	3	0.78	0.15	0.07
	4	0.77	0.04	0.19
	2	0.74	0.11	0.15
	9	0.66	0.28	0.06
	15	0.60	0.37	0.03
	17	0.49	0.20	0.31
$Cl3_{STR}$ South	7	0.02	0.85	0.13
Galician	14	0.04	0.85	0.11
	16	0.27	0.61	0.12
	8	0.03	0.50	0.47
	10	0.22	0.42	0.36
$Cl2_{STR}$ North Spain	12	0.05	0.07	0.88
	11	0.01	0.14	0.85
	13	0.17	0.25	0.58
	5	0.23	0.21	0.56
	6	0.20	0.27	0.53

two northern clusters was the absence or very low frequency of the alleles Sdh^{97}, $Idh\text{-}1^{105}$ and Ugp^{100} in the Mediterranean cluster. Gene flow values between the three identified clusters were: $Nm\ Cl1_{STR}/Cl3_{STR} = 3.08 > Nm\ Cl3_{STR}/Cl2_{STR} = 3.04 > Nm\ Cl2_{STR}\ /\ Cl1_{STR} = 2.75$; thus, we can say that the subpopulations included in $Cl1_{STR}$ are more related to $Cl3_{STR}$ subpopulations than to the northern subpopulations in $Cl2_{STR}$. When *Pgi-2* and *Sdh*, the two loci affected by selection, were removed and analysed with STRUCTURE, the assignment of individuals to the three clusters was maintained, although the percentage of assignment of each individual to the main cluster was reduced.

The D-statistics analysis for all populations revealed values of $D^2_{ST} = 0.071 > D^2_{IS} = 0.001$ and $D'^2_{ST} = 0.0004 < D'^2_{IS} = 0.072$, suggesting that the structure originated mainly from limited migration and drift and not from natural epistatic selection.

The genetic parameters of within populations variability, including *na, ne, Ho, He, Fis, P* and *Ar*, are shown in Table 4. Contributions to the diversity and allelic richness of the subpopulations were included in Figure 4.

The values of *ne* and *He* showed a very high correlation, as was expected (0.97 probability < 0.0001), and only a moderate correlation with *P.* The effective number of alleles (n_e) varied from 1.36 to 1.64 (mean, 1.50). The expected heterozygosity (H_e) varied from 0.22 to 0.35 (mean, 0.28), and the percentage of polymorphic loci (*P*) ranged from 61.54% to 84.62% (mean, 78.28%). There were large differences among populations regarding the effective number of alleles per locus, heterozygosity and levels of polymorphism. The highest values for these genetic parameters corresponded to subpopulations 7, 8, 10, 14, 16 and 17, while the lowest values corresponded to subpopulations 1, 4, 15 and 11. Subpopulations 4 and 7 had the lowest and the highest values, respectively, of both H_e and *P*.

The values of *na, P* and *Ar* showed very high correlations, as was expected (> 0.82, probability < 0.0001). The allelic richness values ranged from 18.72 to 23, and the highest *Ar* values were identified in the two southern Galician subpopulations, 7 and 14, followed by subpopulation 16 (Las Médulas), subpopulations 5 and 6 in the northwestern corner of Galicia and the three hybrid subpopulations, 10 (Asturias), 17 (Guadalupe) and 8 (Meseta Lucense). The lowest allelic richness was identified in the southern subpopulations 4 (Ronda) and in Central Spain subpopulation 1 (El Tiemblo).

Subpopulations 10, 13, 8, 7 and 14, all in the North, displayed the highest levels of heterozygosity. A heterozygote deficit was identified in subpopulations 2 and 4, and a heterozygote excess was identified in the three subpopulations affected by domestication, subpopulations 8, 9 and 15.

The analysis of the contributions of each subpopulation to diversity and allelic richness gave similar results to those based on classical within population diversity parameters described in the precedent paragraphs but added some hidden information (Fig. 4a and 4b). The highest contributions to diversity were made by subpopulations 7 and 14, mainly due to their high own diversity, followed by subpopulations 11 and 12, in this case due to their high divergence. The lowest contributions came from subpopulations 4, 9, 15 and 1 due to their own low diversity. The highest contributions to allelic richness also came from subpopulations 7 and 14 due to their own high diversity and, in the case of population 7, due to its high divergence. The loci with *Rst* < 0 were those with alleles absent or at a low frequency in some stands (< 0.1) (Table 2).

Table 4. Genetic diversity within each stand. Stands were grouped into the clusters obtained with STRUCTURE and ordered within each cluster from highest to lowest ancestry in the cluster. Stands 17 and 10 were considered intermediate between the three clusters

Cluster region	Stand	n_a	n_e	H_o	H_e	*Fis*	*P* (%)	*Ar*
$Cl1_{STR}$ Mediterranean	1	1.77	1.38	0.24	0.23	–0.04	69.23	19.86
	3	1.85	1.45	0.27	0.27	0.00	84.62	20.89
	4	1.69	1.38	0.17	0.22	0.23	61.54	18.72
	2	1.85	1.47	0.23	0.27	0.15	76.92	20.90
	9	1.85	1.42	0.31	0.25	–0.24	84.62	21.79
	15	1.85	1.36	0.26	0.23	–0.13	76.92	21.89
	17	1.92	1.52	0.28	0.30	0.07	84.62	21.93
$Cl3_{STR}$ South Galician	7	2.00	1.64	0.37	0.35	–0.06	84.62	22.87
	14	2.00	1.56	0.33	0.33	0.00	84.62	23.00
	16	2.00	1.56	0.29	0.30	0.03	84.62	22.99
	8	1.92	1.53	0.32	0.30	–0.07	76.92	21.93
	10	1.92	1.61	0.38	0.34	–0.12	76.92	21.99
$Cl2_{STR}$ North Spain	12	1.85	1.50	0.28	0.28	0.00	76.92	20.80
	11	1.77	1.49	0.29	0.27	–0.07	69.23	19.99
	13	1.92	1.56	0.32	0.30	–0.07	76.92	21.98
	5	2.00	1.50	0.28	0.29	0.03	84.62	22.76
	6	1.92	1.50	0.28	0.27	–0.04	76.92	22.00
	Mean	1.89	1.50	0.29	0.28		78.28	21.55

n_a: number of observed alleles per locus. n_e: effective number of alleles per locus. H_o: observed heterozygosity. H_e: expected heterozygosity. *Fis*: fixation index. *P*(%): percentage of polymorphic loci. *Ar*: allelic richness for 11 isozyme loci.

The analysis of F_{st} and N_m at the subpopulation level for the identification of more genetically isolated subpopulations (shown in Table a.3) indicated that when subpopulations 1, 11, 12, 14 and 15 were removed from the calculation, there was a decrease in the mean value of F_{st} and an increase in the mean value of N_m. As mentioned above, this effect indicates a low rate of gene exchange among these subpopulations. The same effect was seen in subpopulations 2, 4 and 7, but with a moderate reduction and increase in the mean values of

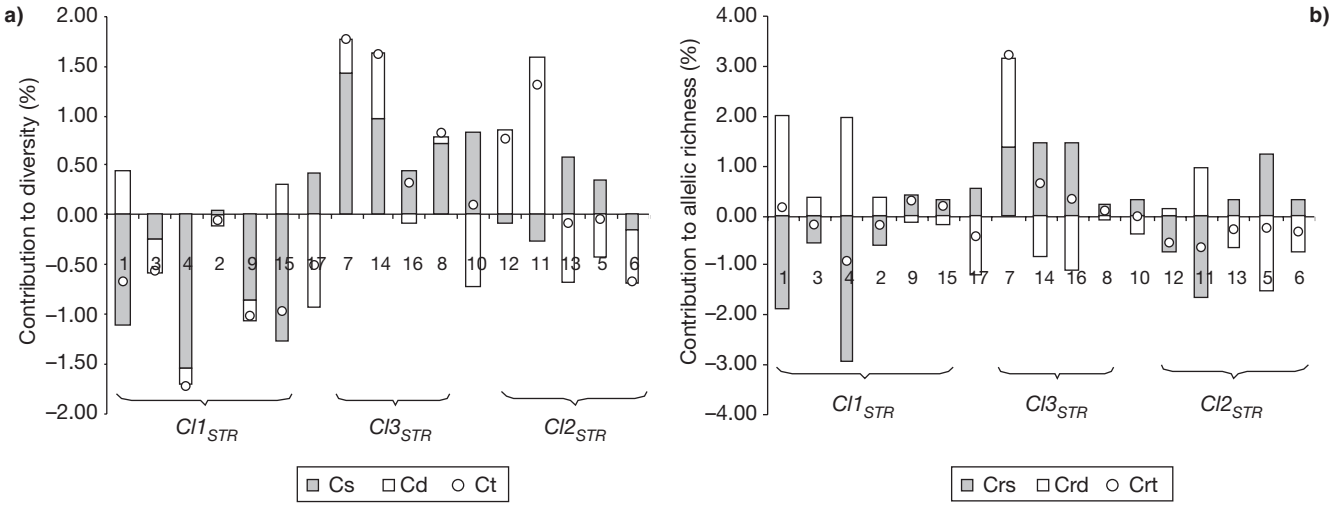

Figure 4. The contribution of each individual *Castanea sativa* stand to global diversity (a) and allelic richness (b). The contribution of each population to the total diversity (*Ct*) and the total allelic richness (*Crt*) was divided in two components: the contribution due to its own diversity and allelic richness (*Cs* and *Crs*) and the contribution due to its own divergence (*Cd* and *Crd*).

F_{st} and N_m, respectively. On the other hand, the opposite effect was seen in subpopulations 5, 6, 10, 13, 17, and especially subpopulation 10, which produced the highest increase in the mean value of F_{st} when removed. These five subpopulations had the highest rates of gene exchange and were more similar to the others. These subpopulations also showed the lowest values of contributions due to differences in diversity and allelic richness (Fig. 4a and 4b).

Discussion

Wild Spanish chestnut populations displayed heterogeneity of allele frequencies and large differences in genetic parameters, indicating high levels of genetic diversity. The mean values of the effective number of alleles per locus, the observed values of heterozygosity and the percentage of polymorphic loci among Spanish populations ($n_e = 1.50$; $H_o = 0.29$; $P = 78.28\%$) were higher than those reported in isozyme studies of the Italian ones ($n_e = 1.41$; $H_o = 0.24$; $P = 60.00\%$; Villani *et al*., 1991 b). The values for the Spanish populations were closer to those obtained for Balkan populations ($H_o = 0.27$; $P = 67.5\%$; Aravanopoulos *et al.,* 2001) and eastern Turkish populations ($n_e = 1.53$; $H_o = 0.28$; $P = 83.75\%$; Villani *et al*., 1994), which had previously displayed the highest levels of genetic variability. The traces of selection in the *Pgi-2* and *Sdh* loci were also identified in Italian populations (Pigliucci, 1990).

The differentiation among populations was also high ($F_{st} = 0.15$) compared with natural populations in Italy ($F_{st} = 0.09$; Villani *et al.,* 1991b) and France ($F_{st} = 0.08$; Bonnefoi, 1984). It was similar to the levels of differentiation previously described for all Turkish populations ($F_{st} = 0.16$; Villani *et al*., 1994), although a little lower.

Clustering methods allowed the identification of a geographic structure of the Spanish populations diversity in three differentiated groups: a Mediterranean group, which includes populations between Andalucía, the central Spanish mountains and Caurel in the northeastern Galician mountains; a northern Spanish group, which comprises the populations of Basque country and the northwestern coast of Galicia; and a southern Galician group, which includes populations in the Pontevedra province, close to northern Portugal, Las Médulas in northwest León, the Meseta Central Lucense and probably northern Portugal. The purest subpopulations within each identified group considering the highest ancestry estimated with STRUCTURE were: the subpopulation from El Tiemblo for the Mediterranean group, the subpopulation from San Cibrán for southern Galician group and the subpopulation from Etxaguen for the northern group. The highest diversity and allelic richness were identified in northern populations, especially in the southern Galician cluster, with the highest allelic richness and heterozygosity. This is the primary gene pool identified in the Iberian chestnut populations, containing all of the isozyme alleles identified.

The clustering method of STRUCTURE is based on linkage and Hardy-Weinberg equilibrium, giving the probability of different numbers of K ancestral populations. In this study, the highest probability was for two and three ancestral populations (data not reported). However, there is also the possibility of one unique ancestral population. The use of additional loci, sampling the different linkage groups, will probably allow a more accurate identification of the number of ancestral populations from which the current populations originated.

Several hypotheses about the origin of the identified Iberian populations gene pools can be established using this set of isozyme loci. The most diverse group of populations (*Cl3*) could have its origin in an Iberian population conserved in a refugia located in northern Portugal and southern Galicia which was previously identified by Krebs *et al.* (2004) using pollen remains. This proposal is based on the fact that the populations close to refugia in temperate regions conserve higher levels of genetic diversity than colonised areas (Hevitt, 1996). Another possibility is a hybrid origin of the southern Galician cluster between the northern population and the Mediterranean populations. This hypothesis is based on the findings in many European tree species that the most diverse populations were located in areas of admixture of different lineages (Petit *et al.,* 2003). Another possibility, the introduction of chestnuts into the Iberian Peninsula from other Mediterranean countries, entering through the South, as has been hypothesised in the literature, is not supported by these data as the highest allelic richness is in northern populations. However, it is necessary to increase the sampling effort in the South Iberian populations.

The presence of a different gene pool in the northern Iberian populations was also detected using intersimple sequence repeat markers (*ISSR*) analysed in populations from the entire European chestnut range (Mattioni *et al*., 2008). In this study, four different European gene pools were identified: two in the Balkan Peninsula; one distributed throughout the rest of the chestnut range,

including Italian, French and most of the Iberian populations; and one in the northern coast of the Iberian Peninsula. The sampling of Spanish populations for the *ISSR* survey included the same individuals from stands 4, 6, 8, 10, 16 and 17 in the present study. In the *ISSR*s study, stands 6 and 10 in northwest Spain were considered to belong to an Atlantic gene pool, different from the main European gene pool, which includes Italian and French populations and Spanish populations represented by stands 4, 8, 16 and 17 (codes for this isozyme study). We also found differentiation of northern Atlantic populations.

The existence of a northern Iberian refuge is supported by the presence of particular haplotypes detected with *RFLP cpDNA* analysis in the Iberian peninsula (Fineschi *et al*., 2000) and in the northwestern Galician and Asturias wild populations (Abuín and Fernández-López, not published). One possibility is that the origin of the two northern demes identified in this isozyme research could be in the two relict populations identified by Krebs *et al.* (2004) based on pollen remains. The hypothesis of the origin of the actual Iberian chestnut populations in relict populations conserved in refugia during the Würm glaciation in the northern Iberian Peninsula is likely considering existing information. It is also supported by evidence for the existence of refugia for other species such as *Quercus* and *Pinus* (Dumolin-Lapègue *et al*., 1997; Olalde *et al*., 2000; Salvador *et al*., 2000; Petit *et al*., 2002).

Spanish populations were frequently considered to be naturalised from introduced germplasm. In case of introduction of selected material by humans, however, it is expected that the geographic structure will show more discontinuous patches originating from the translation of genetic resources by man, and not a structure with a continuous pattern of variation originating from gene flow as was identified in this analysis.

It is interesting to note that the identified geographic structure precedes the introduction of Asiatic germplasm. The Japanese chestnut and some Chinese chestnuts were introduced to northern Spain from 1920 until 1940, and since 1960, inter-specific hybrids, mainly F_1 and F_2 *C. crenata x C. sativa*, have been used in plantations, predominantly in Galicia. Diagnostic and specific alleles alleles of the Asiatic species were not present in the gels; furthermore, some of the exclusive alleles of $Pgi\text{-}2^{105}$ and $Mdh\text{-}3^{95}$ in northern populations were not present in the Asiatic species (Fernández-López, 1996). Thus, the identified variability is specific for the sweet chestnut.

The clusters obtained with isozymes were quite similar to the grouping obtained in a previous study of Spanish chestnut populations using flushing time, an important adaptive trait in temperate regions (Fernández-López *et al*., 2005). The aforementioned study indicated a pattern of structured variability caused by natural selection, as was demonstrated by the correlation of flushing time with drought. Mediterranean populations showed an early flushing time and very low additive genetic variation within populations, while northern populations showed a late flushing time and the highest additive variance within. Thus, the natural selection is another evolutionary force acting over the different identified gene pools, increasing the adaptation of populations to local conditions.

An understanding of the genetic diversity of wild chestnut populations is one of the starting points for developing a conservation and management strategy. According to the European Forest Gene Conservation programme, a network of *in situ* populations must be selected for conservation. The selection of Conservation Units must consider the inclusion of representatives from different extant evolutive units. We propose the conservation of the three identified clusters through the conservation of representative stands considering their conservation value (*e.g.,* difference, diversity and allelic richness). According to the results of this study some of the stands are very interesting due to their contribution to diversity and allelic richness could present difficulties for conservation management because they are located on private property (*e.g.,* stand 7 in San Cibrán). As a representative of the southern Galician deme, we propose the Fraga de Catasós stand (stand 14), which is a protected area. A good representative of the northern deme would be Fragas del Eume (stand 13), a Natural Park, and Hervás and El Tiemblo (stands 2, 3 and 1) should be considered for the Mediterranean populations. Populations that had not been studied in this study in the northern coast and in Andalucía could also be interesting. A network of seed stands and seed sources for seed collection to be used in plantations in the same region is a form of dynamic conservation which includes adaptation.

Acknowledgements

This study was financially supported by the European Union research project of EVK2-CT-1999-00006, by 92-*CastaneaREG* INTERREG III B «Espacio Atlanti-

co» and the complementary grant PGIDIT04PXIC50201P (IN.CI.TE. Xunta de Galicia). The authors thank Raquel Blanco Silva for her field and laboratory work, Taciana Rial for laboratory assistance and two anonymous referees by their careful revisions. Special thanks go to Elena Vilches Plaza from Diputación de Alava and to technicians from Lourizán (Agustín Quintairós, Miguel Jamardo and César) for sample collection from the Antoñana and Etxaguen populations.

References

ARAVANOPOULOS F.A., BUCCI G., AKKAK A., BLANCO SILVA R., BOTTA R., BUCK E., CHERUBINI M., DROUZAS A.D., FERNÁNDEZ-LÓPEZ J., MATTIONI C., MARINONI D., PAPADIMA A., RUSSELL K., ZAS R., VILLANI F., 2005. Molecular population genetics and dynamics of chestnut (*Castanea sativa*) in Europe: inferences for gene conservation and tree improvement. Acta Horticulturae 693, 403-412.

ARAVANOPOULOS F.A., DROUZAS A.D., ALIZOTI P.G., 2001. Electrophoretic and quantitative variation in chestnut (*Castanea sativa* Mill.) in Hellenic Forest Snow & Landscape Research 76(3), 429-434.

BENITO-GARZÓN M., SÁNCHEZ DE DIOS R., SAINZ-OLLERO H., 2007. Predictive modelling of tree species distribution on the Iberian Peninsula during the Last Glacial Maximum and Mid-Holocene. Ecography 30, 120-134.

BONNEFOI C., 1984. Etude du polymorphisme enzymatique des populations forestières de châtaignier, *Castanea sativa* Miller. Thèse de doctorat. Univ des Sciences et Techniques du Languedoc, Académie de Montpellier.

CACCONE A., SBORDONI V., 1987. Molecular evolutionary divergence among North American cave crickets. I. Allozyme variation. Evolution 41(6), 1198-1214.

CARRIÓN J.S., YIL E.I., WALKER M.J., LEGAZ A.J., CHAÍNS C., LÓPEZ, A., 2003. Glacial refugia of temperate, Mediterranean and Ibero-North African flora in southeastern Spain: new evidence from cave pollen at two Neanderthal man sites. Global Ecol Biogeog 12, 119-129.

CONEDERA M., KREBS P., TINNER W., PRADELLA M., TORRIANI D., 2004. The cultivation of *Castanea sativa* (Mill.) in Europe, from its origin to its diffusion on a continental scale. Veget Hist Archaeobot 13, 161-179.

DUMOLIN-LAPÈGUE S., DEMESURE B., FINESCHI S., LE CORRE V., PETIT R.J., 1997. Phylogeographic structure of white oak throughout the European continent. Genetics 146, 1475-1487.

FERNÁNDEZ-LÓPEZ J., 1996. Variabilidad isoenzimática, morfológica y selección clonal en *Castanea sativa* Miller, *Castanea crenata* Sieb. et Zucc. *Castanea mollissima* e híbridos interespecíficos. Ph D thesis. Univ Politécnica de Madrid.

FERNÁNDEZ-LÓPEZ J., ZAS R., BLANCO-SILVA R., DÍAZ R., 2005. Geographic differentiation in adaptative traits of wild chestnut Spanish populations (*Castanea sativa* Miller). Invest Agrar: Sist Recur For 14(1), 13-26.

FINESCHI S., TAURCHINI D., VILLANI F., VENDRAMIN G.G., 2000. Chloroplast DNA polymorphism reveals little geographical structure in *Castanea sativa* Mill. (Fagaceae) throughout southern European countries. Mol Ecol 9, 1495-1503.

HARTL D.L., CLARK A.G. (eds), 1989. Principles of population genetics, 2nd ed. Sinauer Associates, Sunderland, MA.

HEWITT G.M., 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. Biol J Linn Soc 58, 247-276.

HUANG H.W., DANE F., NORTON J.D., 1994. Genetic analysis of 11 polymorphic isozyme loci in chestnut species and characterization of chestnut cultivars by multi-locus allozyme genotypes. J Am Soc Hort Sci 119(4), 840-849.

HUNTLEY B., BIRKS H.J.B. (eds), 1983. An atlas of past and present pollen maps for Europe: 0-13,000 years ago. Cambridge University Press, Cambridge.

JACKOBSSON M., ROSENBERG.N., 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23, 1801-1806.

KIM Z.S., 1979. Inheritance of leucine aminopeptidase and acid phosphatase isozymes in beech (*Fagus sylvatica* L.). Silvae Genet 28(2-3), 68-71.

KREBS P., CONEDERA M., PRADELLA M., TORRIANI D., FELBER M., TINNER W., 2004. Quaternary refugia of the sweet chestnut (*Castanea sativa* Mill.): an extended palynological approach. Veget Hist Archaeobot 13, 145-160.

LEVENE H., 1949. On a matching problem in genetics. Ann Math Stat 20, 91-94.

MACHON N., BUREL L., LEFRANC M., FRASCARIA N., 1996. Evidence of genetic drift in chestnut populations. Can J For Res 26(5), 905-908.

MALVOLTI M.E., FINESCHI S., 1987. Analysis of enzyme systems in chestnut (*Castanea sativa* Mill.). Genet Agrar 41(3), 243-255.

MANLY B.F.J. (ed), 1985. The statistics of natural selection. Chapman and Hall, London.

MARTÍN-ARROYO T., RUIZ-ZAPATA M.B., PÉREZ-GONZÁLEZ A., VALDEOLMILLOS A., DORADO VALIÑO M., BENITO G., GIL GARCÍA M.J., 1999. Paleoclima y paleoambiente durante el pleistoceno superior y el tardiglaciar en la region Central Peninsular. In: Avances en el estudio del cuaternario español (Palli Buxó L., Roqué Pau C., eds). Girona. pp. 317-324.

MATTIONI C., CHERUBINI M., MICHELI E., VILLANI F., BUCCI G., 2008. Role of domestication in shaping *Castanea sativa* genetic variation in Europe. Tree Genetics and Genomes 4, 563-574.

NEI M., 1973. Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci 70, 3321-3323.

NEI M., 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89, 583-590.

NEI M., 1987. Molecular Evolutionary Genetics. Columbia Univ Press, NY.

OHTA T., 1982a. Linkage disequilibrium with the island model. Genetics 101, 139-155.

OHTA T., 1982b. Linkage disequilibrium due to random drift in finite subdivided populations. Proc Natl Acad Sci 79, 1940-1944.

OLALDE M., HERRÁN A., ESPINEL S., GOICOECHEA P.G., 2002. White oaks phylogeography in the Iberian Peninsula. For Ecol Manag 156, 89-102.

PETIT R.J., EL MOUSADIK A., PONS O. 1997. Identifying populations for conservation on the basis of genetic markers. Conservation Biology 12(4), 844-855.

PETIT R.J., BREWER S., BORDACS S., BURG K., CHEDDADI R., COART E., COTTRELL J., CSAIKL U.M., VAN DAM B., DEANS J.D., ESPINEL S., FINESCHI S., FINKELDEY R., GRAZ I., GOICOECHEA P.G., JENSEN J.S., KÖNIG A.O., LOWE A.J., MADSEN S.F., MÁTTYÁS G., MUNRO R.C., POPESCU F., SLADE D., TABBENER H., DE VRIES S.G.M., ZIEGENHAGEN B., BEAULIER J.L., KREMER A., 2002. Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. For Ecol Manag 156, 49-74.

PETIT R.J., AGUINAGALDE I., DE BEAULIEU J.L., BITTAKAU C., BREWER S., CHEDDADI R., ENNOS R., FINESCHI S., GRIVET D., LASCOUX M., MOHANTY A., MÜLLER-STARCK G., DEMESURE-MUSCH B., PALMÉ A., MARTÍN J.P., RENDELL S., VENDRAMIN G.G., 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. Science, 300, 1563-1565.

PIGLIUCCI M., BENEDETTELLI S., VILLANI F., 1990. Spatial patterns of genetic variability in Italian chestnut (*Castanea sativa*). Can J Bot 68, 1962-1967.

PRITCHARD J.K., STEPHENS M., DONNELLY P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945-959.

ROHLF F.J., 2000. NTSYSpc. Numerical taxonomy and multivariate analysis system v. 2.1. Exeter Software, Setauket, NY.

ROSENBERG N.A., 2004. DISTRUCT: a programme for the graphical display of populations structure. Molecular Ecology Notes 4, 137-138

SALVADOR L., ALÍA R., AGÚNDEZ D., GIL L., 2000. Genetic variation and migration pathways of maritime pine (*Pinus pinaster* Ait.) in the Iberian Peninsula. TAG 100, 89-95.

SHIELDS C.R., ORTON T.J., STUBER C.W., 1983. An outline of general resource needs and procedures for electrophoretic separation of active enzymes from plant tissue. In: Isozymes in Plant genetics and Breeding. Part A (Tanksley S.D., Orton T.J., eds). Elsevier Science Publishers B, Amsterdam. pp. 425-468.

SLATKIN M., BARTON N.H., 1989. A comparison of three indirect methods for estimating average levels of gene flow. Evolution 43, 1349-1368.

VILLANI F., BENEDETTELLI S., PACIUCCI M., CHERUBINI M., PIGLIUCCI M., 1991a. Genetic variation and differentiation between natural populations of chestnut (*Castanea sativa* Mill.) from Italy. In: Biochemical markers in the population genetics of forest trees (Hattemer H.H., Fineschi S., eds). SPB Academic Publishing, The Hague, The Netherlands. pp. 91-103.

VILLANI F., PIGLIUCCI M., BENEDETTELLI S., CHERUBINI M., 1991b. Genetic differentiation among Turkish chestnut (*Castanea sativa* Mill.) populations. Heredity 66, 131-136.

VILLANI F., PIGLIUCCI M., CHERUBINI M., 1994. Evolution of *Castanea sativa* Mill. in Turkey and Europe. Genet Res Camb 63, 109-116.

WEIR B.S. (ed), 1990. Genetic data analysis. Sinauer Associates, Sunderland, MA.

WENDEL J.F., WEEDEN N.F., 1989 Visualization and interpretation of plant isozymes. In: Isozymes in plant biology (Soltis D.E., Soltis P.E., eds). Discorides Press, Portland, Oregon. pp. 5-45.

WRIGHT S. (ed), 1978. Variability within and among natural populations. Univ of Chicago Press, Chicago. Vol 4.

YEH F.C., BOYLE T., 1997. Popgene v. 1.31. Microsoft Window-based freeware for population genetic analysis. Univ of Alberta and Center for International Forestry Research, Alberta, Canada.

ZOHARY D., HOPF M. (eds), 1988. Domestication of plants in the old world. Clarendon, Oxford.

Additional Table A.1. Allelic frequencies of the polymorphic loci calculated for each stand and number of alleles per stand. Stands were grouped according to their clustering with STRUCTURE

Locus	Allele	$CI1_{STR}$ Mediterranean							$CI3_{STR}$ Southern Galician					$CI2_{STR}$ North Iberian				
		1	3	4	2	9	15	17	7	14	16	8	10	12	11	13	5	6
Pgi-2	108	0.47	0.47	0.20	0.62	0.47	0.50	0.43	0.52	0.68	0.63	0.17	0.38	0.33	0.08	0.17	0.23	0.22
	105	*0.02*					*0.02*		**0.45**	**0.08**	**0.12**	**0.19**	**0.05**			*0.05*	*0.02*	*0.09*
	100	0.52	0.53	0.80	0.38	0.53	0.48	0.57	0.03	0.23	0.25	0.64	0.57	0.67	0.92	0.78	0.75	0.69
Sdh	100	0.92	0.95	0.35	0.83	0.60	0.78	0.31	0.58	0.65	0.42	0.40	0.58	0.17	0.18	0.55	0.73	0.54
	97			*0.03*	*0.13*			*0.10*	**0.20**	**0.12**	**0.17**	**0.57**	**0.08**	*0.52*	*0.42*	*0.22*	*0.12*	*0.16*
	95	0.08	0.05	0.62	0.03	0.40	0.22	0.59	0.22	0.23	0.42	0.03	0.33	0.32	0.40	0.23	0.15	0.30
Idh-1	108	*0.08*	*0.08*		*0.13*	*0.12*	*0.07*	*0.17*	**0.45**	**0.77**	**0.23**	**0.19**	**0.37**	*0.73*	*0.80*	*0.45*	*0.55*	*0.48*
	100	0.92	0.92	1.00	0.87	0.88	0.93	0.83	0.55	0.23	0.77	0.81	0.63	0.27	0.20	0.55	0.45	0.52
Idh-2	105	*0.27*	*0.15*	*0.03*	*0.20*	*0.05*	*0.09*	*0.07*	**0.12**	**0.17**	**0.07**	**0.03**				*0.05*	*0.02*	*0.04*
	100	0.73	0.85	0.97	0.80	0.95	0.92	0.93	0.88	0.83	0.93	0.97	1.00	1.00	1.00	0.95	0.98	0.96
Ugp-1	100		*0.08*		*0.02*	*0.02*		*0.14*	**0.17**	**0.15**	**0.08**	**0.24**	**0.18**	*0.23*	*0.27*	*0.23*	*0.25*	*0.17*
	95	1.00	0.92	1.00	0.98	0.98	1.00	0.86	0.83	0.85	0.92	0.76	0.82	0.77	0.73	0.77	0.75	0.83
Mdh-2	105	0.35	0.43	0.31	0.62	0.03	0.08	0.34	0.15	0.18	0.13	0.45	0.33	0.20	0.22	0.12	0.30	0.17
	100	0.65	0.57	0.69	0.38	0.97	0.92	0.66	0.85	0.82	0.87	0.55	0.67	0.80	0.78	0.88	0.70	0.83
Mdh-3	100	1.00	0.97	1.00	1.00	0.77	0.83	0.97	0.70	0.85	0.88	1.00	0.87	0.98	1.00	1.00	0.93	1.00
	95		*0.03*			*0.23*	*0.17*	*0.03*	**0.30**	**0.15**	**0.12**		**0.13**	*0.02*			*0.07*	
Adh	100	0.81	0.55	0.69	0.57	0.73	0.88	0.71	0.48	0.65	0.50	0.71	0.43	0.73	0.45	0.58	0.35	0.52
	95	0.19	0.45	0.31	0.43	0.27	0.12	0.29	0.52	0.35	0.50	0.29	0.57	0.27	0.55	0.42	0.65	0.48
Prx-2	105	0.58	0.37	0.35	0.42	0.43	0.38	0.34	0.13	0.47	0.40	0.22	0.48	0.45	0.43	0.53	0.30	0.06
	100	0.42	0.63	0.65	0.58	0.57	0.62	0.66	0.87	0.53	0.60	0.78	0.52	0.55	0.57	0.47	0.70	0.94
Dia-1	110	0.82	0.67	0.58	0.48	0.25	0.12	0.38	0.38	0.27	0.10	0.74	0.28	0.65	0.85	0.53	0.42	0.48
	100	0.18	0.33	0.42	0.52	0.75	0.88	0.62	0.62	0.73	0.90	0.26	0.72	0.35	0.15	0.47	0.58	0.52
Dia-2	110	0.92	0.82	0.92	0.90	0.85	0.67	0.84	0.53	0.37	0.58	0.67	0.73	0.98	0.77	0.83	0.93	0.90
	100		*0.18*	*0.08*	*0.08*	*0.15*	*0.33*	*0.10*	**0.47**	**0.63**	**0.42**	**0.33**	*0.16*	*0.02*	*0.23*	*0.17*	*0.07*	*0.10*
Nº alleles		20	22	20	22	22	22	23	24	24	24	23	23	22	21	23	24	23

Additional Table A.2. Nei's (1978) unbiased genetic distances among wild chestnut stands

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	—																
2	0.0291	—															
3	0.0152	0.0102	—														
4	0.0565	0.0638	0.0485	—													
5	0.0927	0.0610	0.0499	0.0714	—												
6	0.0942	0.0744	0.0575	0.0484	0.0140	—											
7	0.1391	0.0981	0.0921	0.1299	0.0794	0.0602	—										
8	0.0721	0.0656	0.0517	0.0563	0.0700	0.0501	0.0956	—									
9	0.0674	0.0644	0.0557	0.0378	0.0662	0.0541	0.0728	0.0983	—								
10	0.0807	0.0440	0.0434	0.0491	0.0182	0.0334	0.0557	0.0699	0.0292	—							
11	0.1528	0.1649	0.1377	0.1048	0.0592	0.0572	0.1509	0.0766	0.1521	0.0865	—						
12	0.1162	0.1200	0.1177	0.0847	0.0556	0.0454	0.1224	0.0647	0.0988	0.0690	0.0230	—					
13	0.0631	0.0729	0.0528	0.0450	0.0197	0.0244	0.0911	0.0494	0.0449	0.0226	0.0383	0.0292	—				
14	0.1462	0.1144	0.1195	0.1652	0.0895	0.0934	0.0447	0.1405	0.0878	0.0551	0.1349	0.1053	0.0824	—			
15	0.0764	0.0664	0.0618	0.0666	0.0927	0.0783	0.0763	0.1074	0.0097	0.0463	0.2070	0.1407	0.0710	0.0793	—		
16	0.1230	0.0754	0.0846	0.0791	0.0772	0.0675	0.0376	0.1089	0.0267	0.0236	0.1610	0.1159	0.0689	0.0477	0.0314	—	
17	0.0667	0.0482	0.0499	0.0122	0.0545	0.0359	0.0823	0.0545	0.0227	0.0232	0.1009	0.0625	0.0377	0.0982	0.0427	0.0354	—

Additional Table A.3. Values of gene flow and degree of genetic differentiation after the exclusion of each stand in turn

Population excluded	*Fst*	*Nm*
1	0.147	1.455
2	0.151	1.401
3	0.153	1.385
4	0.150	1.415
5	0.155	1.365
6	0.155	1.362
7	0.151	1.410
8	0.152	1.398
9	0.153	1.382
10	0.159	1.321
11	0.143	1.501
12	0.149	1.430
13	0.157	1.340
14	0.148	1.442
15	0.148	1.434
16	0.153	1.389
17	0.157	1.343
Average value	0.152	1.392

Additional Table A.4. *Fst* values among stands

Pop.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	—																
2	0.046	—															
3	0.027	0.018	—														
4	0.093	0.093	0.076	—													
5	0.118	0.075	0.064	0.095	—												
6	0.125	0.093	0.075	0.071	0.022	—											
7	0.147	0.101	0.097	0.140	0.081	0.068	—										
8	0.096	0.079	0.152	0.080	0.078	0.062	0.088	—									
9	0.098	0.086	0.076	0.061	0.084	0.074	0.082	0.114	—								
10	0.098	0.053	0.053	0.066	0.024	0.043	0.054	0.072	0.040	—							
11	0.183	0.176	0.155	0.136	0.074	0.075	0.142	0.090	0.172	0.093	—						
12	0.146	0.136	0.135	0.112	0.069	0.061	0.119	0.077	0.121	0.076	0.034	—					
13	0.085	0.086	0.066	0.063	0.027	0.035	0.088	0.058	0.059	0.028	0.050	0.039	—				
14	0.157	0.117	0.123	0.172	0.092	0.101	0.046	0.129	0.099	0.056	0.135	0.109	0.084	—			
15	0.117	0.094	0.089	0.107	0.118	0.108	0.092	0.131	0.020	0.063	0.063	0.170	0.094	0.098	—		
16	0.145	0.088	0.098	0.104	0.086	0.081	0.042	0.112	0.038	0.029	0.163	0.124	0.076	0.053	0.047	—	
17	0.089	0.061	0.063	0.022	0.065	0.048	0.082	0.063	0.034	0.029	0.114	0.075	0.046	0.098	0.062	0.044	—
Total	0.152																