# Landscape genetic structure of chestnut (*Castanea sativa* Mill.) in Spain

M.A. Martín[1*], C. Mattioni[2], J.R. Molina[3], J.B. Alvarez[1], M. Cherubini[2], M.A. Herrera[3], F. Villani [2], L.M. Martín[1]

1) Departamento de Genética, Escuela Técnica Superior de Ingenieros Agrónomos y de Montes, Edificio Gregor Mendel, Campus de Rabanales, Universidad de Córdoba, ES-14071 Córdoba, Spain. *Corresponding author: ge2macum@uco.es

2) Istituto di Biologia Agroambientale e Forestale (IBAF), Consiglio Nazionale delle Ricerche (CNR), Viale Marconi, 2, 05010 Porano (TR), Italy.

3) Departamento de Ingeniería Forestal, Escuela Técnica Superior de Ingenieros Agrónomos y de Montes, Edificio Leonardo Da Vinci, Universidad de Córdoba, ES-14071 Córdoba, Spain.

**For:** Tree Genetics & Genomes

## Abstract

The current need for forest conservation and management has driven a rapid expansion of landscape genetics approach. This discipline combines tools from molecular genetics, landscape ecology and spatial statistics and is decisive for improving not only ecological knowledge but also for properly managing population genetic resources. This approach could be appropriate to sweet chestnut (*Castanea sativa* Mill.), a multipurpose species of great economic importance in the Mediterranean basin, and a species considered to be a good model of integration between natural and human-driven distribution of diversity. Sixteen chestnut populations, covering the distribution range of the species in Spain, were analysed using seven microsatellite markers. Results revealed a high level of genetic diversity in Spanish chestnut populations, which in part followed a geographical pattern, although distribution was not homogeneous. Likewise, areas particularly rich in diversity were detected, facilitating the development of a hypothesis about the history of chestnut in Spain. In conclusion, these results provide valuable baseline data for more in-depth studies on chestnut landscape genetics that can contribute to its conservation.

**Keywords:** sweet chestnut, genetic diversity, landscape genetics

## Introduction

Forests are key ecosystems with high levels of biodiversity that supply a multitude of environmental, economic and aesthetic benefits (Pautasso 2009). An important indicator of biodiversity is the amount of genetic diversity provided by the dominant tree species. Trees are ecosystem engineers and landscape modulators in the sense that they create resource niches for a whole suite of other organisms dependent on the development, structural support, decay and renewal of trees (Wright and Jones 2006; Shackak et al. 2008).

The persistence of populations is positively linked to their genetic variability (Frankham and Ralls 1998; Saccheri et al. 1998), which is widely considered the main requirement for the long-term survival of species on an evolutionary time scale (Booy et al. 2000; Namkoong 2001). Thus, genetic diversity and population genetic structure are key elements that determine the ability of a species to respond to selection, either natural or artificial, and its distribution over the range of a species is the result of environmental and management factors that have shaped its evolution (Fady and Conord 2010).

Landscape features (elevation, slope, etc.) can influence genetic structuring of populations at a regional level because they can affect gene flow (Gomez et al. 2005). Studies on habitat fragmentation, connectivity of populations and human impact have been carried out to understand how landscape shapes the genetic structure of species (Petit et al. 2002, Oddou-Muratorio et al. 2004, Sork and Smouse 2006). This information critical for improving both the ecological knowledge and the managment of populations genetic resources (Storfer et al. 2007; Holderegger and Wagner 2008). The current need for increased understanding of these processes has driven a rapid expansion of landscape genetics approach. This discipline combines tools from molecular genetics, landscape ecology and spatial statistics, enabling the spatial

mapping of allele frequencies from a species or a group of populations and, subsequently, the correlation of such patterns with the current landscape (Manel et al. 2003; Latta 2006). In this respect, this approach should be appropriate for the analysis of chestnut that is considered a good model of integration between natural and human-driven distribution of diversity.

Tools to characterise the DNA variation of long–lived species are being widely used since they can greatly facilitate prioritization in conservation decisions (Allendorf et al. 2010). Powerful molecular genetic techniques can aid in the management of populations by confirming the identity of accessions and by monitoring forest genetic resources. In particular, microsatellite markers (SSRs) have become the marker of choice for several reasons including their particular structure, high polymorphism rate, codominant nature and wide distribution across the genomes of higher organisms (Tautz and Renz 1984; Gupta et al. 1996; Powell et al. 1996). Microsatellites are excellent markers in population genetic diversity studies for assessing gene flow, effective population size, migration and dispersal processes, parentage and relatedness (Vendramin et al. 2004).

Sweet chestnut (*Castanea sativa* Mill.), the only native species of *Castanea* genus in Europe, is distributed in the majority of the Mediterranean countries, extending from Caucasus to Italy, France, Spain, Portugal and South England. Thus, this wide-range distribution throughout southern Europe highlights the ability of the species to adapt to varying environmental conditions (Lauteri et al. 1998; Martín et al. 2010). Chestnut is one of the multipurpose species of major economic importance in the Mediterranean basin, valued not only for fruit and timber but also for its contribution to the landscape and environment. Because of the multipurpose characteristics of the species, chestnut populations have been affected by clonal propagation, silvicultural practices, etc. This along with the changes in land use and the accelerating dynamics of global and climatic

changes have resulted in a fragmentation of habitats, a reduction of population size and probably, a loss in biodiversity.

The genetic structure of chestnut across Europe was assessed using different molecular markers: isozymes, chloroplast DNA, ISSR, SSR and EST-SSR (Villani et al. 1999; Fineschi et al. 2000; Mattioni et al. 2008; Martin et al. 2010). These studies highlighted both high genetic diversity located in different gene pools along the southern parts of the continent and significant differences among populations in relation to the local adaptation of the species. However, knowledge of the genetic structure of this species in Spain is limited and fragmentary. Most of the existing information is derived from studies of chestnut varieties that come from clonal propagation and are highly managed by humans. These studies have stated a large genetic diversity in these traditional varieties (Martín et al. 2009; Pereira-Lorenzo et al. 2010). With respect to natural chestnut populations, there are a few studies on the genetic structure of the species using ISSR and isozyme markers (Mattioni et al. 2008; Fernandez-Lopez and Monteagudo 2010); however, this information is incomplete because the material used is not representative of the distribution range of the species in Spain.

The aim of this study was to use microsatellite markers to assess the patterns of genetic structure and diversity of natural chestnut populations in Spain in order to: 1) inventory the genetic resources of the species, 2) verify areas as reservoirs of genetic diversity, and 3) discuss the implications of the results for the identification of conservation and management units.

## Material and Methods

*Plant material*

In Spain, chestnut has a discontinuous distribution, with the largest areas located in the north (Asturias, Galicia, Navarra, Cataluña) and scattered stands in the centre and south (Andalusia). It can be found from sea level to 1800 m in altitude and needs annual rainfall between 500–1000 mm (Rivas-Martínez 1987). In general, it grows over broad span climatic conditions from wet and cool to hot and dry during the growth periods. Chestnut occupies three climatic subregions (Atlantic, Central European and Mediterranean) and the main limitation to chestnut diffusion is soil type, because it cannot withstand high levels of active calcium (Allué 1990).

Sixteen chestnut populations (high forest) were analysed, covering the distribution of the species in Spain (Fig. 1). The collection sites were located in climatologically and phytogeographically different areas in the Iberian Peninsula. The detailed location of populations and the number of trees analysed are shown in Table 1.

*DNA extraction and SSR analysis*

Genomic DNA was extracted from 20 mg of lyophilised leaves and fresh buds using the Qiagen DNeasy$^{TM}$ Plant mini Kit protocol.

Seven genomic microsatellites developed in *C. sativa* (Marinoni et al. 2003; Buck et al. 2003) were evaluated. PCR reaction mixture consisted of 20 μL total volume containing 20 ng of genomic DNA following the Qiagen multiplex kit protocol. Cycling parameters were: 15 min at 95°C, 30 cycles of 30 s at 94°C, 90 s at 57°C, and 1 min at 72°C, and a final step of 30 min at 72°C. Amplification products (0.1−1 μL) were added to 20 μL formamide and 0.3 μL Genescan-500 ROX and denatured at 95°C for 5 min and run on an ABI PRISM 3100 DNA sequencer. Allele scoring was performed using the GeneScan 3.5 and Genotyper 3.7 softwares (Applied Biosystems).

*Statistical analyses*

A set of measures of intra and inter population genetic statistics were calculated using Popgene 3.2 (Yeh et al. 1997): number of alleles per locus (A), effective number of alleles (Ne), observed heterozygosity (Ho), expected heterozygosity (He) and deviation from Hardy-Weinberg. Allelic richness (Ar) based on a minimum sample size of 11 individuals was calculated using FSTAT (Goudet 2002). Deficits in heterozygotes that could be attributed to the presence of null alleles, long alleles dropout or scoring errors due to stuttering in the SSRs profile was tested with Micro-checker 2.2.3 (Van Oosterhout et al. 2004). The Ewens-Watterson test for neutrality was performed for each locus to detect possible effects of selection on the distribution of alleles in each population (Manly 1985).

The fixation index $F_{IS}$ (Weir and Cockerham 1984) was calculated using Arlequin 3.1 software (Schneider et al. 2000) and its deviation from zero tested by 10,000 allele permutations. Population differentiation ($F_{ST}$ according to Weir and Cockerham 1984) and standardised $F'_{ST}$ (Hedrick 2005) were calculated using FSTAT software.

In order to detect changes in diversity measures within the assessed material, Pearson's correlations (r) were calculated for Ar, He, $F_{IS}$ and geographical parameters (latitude, longitude and altitude).

Different approaches were used to characterise the pattern of spatial distribution of genetic variation of the samples. The correlation between pairwise genetic distances and pairwise geographic coordinates was tested through a Mantel test using NTSYS 3.2 software (Exeter Software, Setauker, NY USA). For the principal coordinates analysis (PCA) a dissimilarity matrix based on Nei's genetic distance (Nei 1972) was performed with NTSYS 3.2 software. The population structure was performed using a model-based Bayesian approach implemented in software STRUCTURE v.2.1 using the admixture model (Pritchard et al. 2000). This attempts to reveal the population structure by placing

individuals or predefined groups in *K* number of clusters. In this study, the range of possible number of clusters (*K*) tested was 1-19 (the putative number of populations plus 3). Based on the initial results a series of six independent runs were performed for *K* between 1 and 6 with a burn in period of 10000 steps followed by $10^5$ MCMC replicates. To identify the number of clusters (*K*) that best explained the data, the rate of change on L(*K*) ($\Delta K$) between successive *K* values was calculated according to Evanno et al. (2005). The six runs for each simulation were averaged using CLUMPP software (Jakobsson and Rosenberg 2007) and represented graphically with DISTRUCT (Rosenberg 2004). Furthermore, the genetic landscape shape analysis to identify possible genetic discontinuities and barriers was performed using Monmonier's maximum difference algorithm (Monmonier 1973). This algorithm finds the edges associated with the highest rate of change in a given distance measure and is applied to a geometric network that connects all populations using Delaunay triangulation (Watson 1992). The analysis was conducted using Alleles in the Space (AIS) software (Miller 2005). Averages interindividual genetic distances were calculated between populations connected in the network. Next, a simple interpolation procedure was used to infer genetic distances at locations on a uniformly spaced grid (X and Y coordinates correspond to geographical location and Z to genetic distances), to obtain a graphical representation of chestnut genetic patterns across the entire sampled landscape.

## Results

*Genetic diversity*

The seven SSR loci were highly polymorphic and the number of detected alleles for each locus was between 6 (locus *EMCs25*) and 23 (locus *CsCAT3*) (Table 2). A total of 92 different alleles was identified in the 239 individuals with a mean of 13.14 alleles per

locus (Table 2). The gene diversity was high overall (mean = 0.804), varying from 0.646 for locus *CsCAT1* to 0.898 for locus *EMCs38* (Table 2). In general, allele frequencies were distributed unevenly within the loci evaluated. Only 6 alleles presented frequencies in the upper 20%, 38 alleles between 5 and 20%, and 36 were catalogued as rare alleles with a frequency below 5%. All loci showed rare alleles, and 15 private alleles were identified in three of the seven loci (CsCAT1, CsCAT3 and CsCAT6).

The fixation index (measure of an excess of homozygosity) showed positive and significant deviation from zero in only one locus ($F_{IS}$ = 0.431) (Table 2). The results of the Ewens-Watterson neutrality test did not find loci under balancing selection, indicating that the seven loci were neutral (results not shown). Likewise, the excess of homozygotes in locus *EMCs25* may derive from the presence of null alleles in the assessed material, which was detected by Micro-checker.

Genetic diversity parameters for each population based on allelic frequencies are presented in Table 3. Within the 16 populations, the allelic richness over the seven loci for each population ranged from 3.97 in SP10 to 5.82 in SP03 and SP07. The highest mean expected heterozygosity (He) was found in SP14, whereas the lowest value of heterozygosity was detected in SP01 (Table 3). All populations contained rare alleles and the 15 private alleles were distributed over 9 out of the 16 populations: SP01 (1), SP03 (3), SP05 (1), SP07 (2), SP09 (1), SP12 (1), SP14 (2), SP15 (3), SP16 (1). Moreover, when northern and southern populations were grouped, 10 private alleles were detected in northern populations and 13 in southern populations.

The fixation index deviated significantly from zero in only two populations SP04 and SP09, indicating that the rest of the populations were in Hardy-Weinberg equilibrium. Considering that estimated fixation values can be affected by the presence of null alleles

these parameters were recalculated using the six loci that did not show biases. Thus, corrected $F_{IS}$ values displayed similar values as those obtained using the complete set of markers, indicating that with the six loci populations SP04 and SP09 again deviated significantly from zero (Table 3).

*Population structure*

The coefficient of genetic differentiation among the 16 populations ($F_{ST}$) was 0.152, whereas the standardised $F'_{ST}$ was slightly lower, 0.141 (Table 3). Pairwise $F_{ST}$ values ranged from 0.021 between populations SP05 and SP09 to 0.158 between SP09 and SP13 (Supplementary Table).

A Mantel test was performed to examine the correlation between genetic differentiation and geographical distance for the 16 populations. The results revealed a non-significant correlation between both variables ($r = 0.698$, $p = 0.383$). However, principal coordinate analysis suggested the presence of a nonrandom association of populations (Fig. 2). The combined first two axes explained 47.12% of the variation. Results revealed geographical trends of differentiation between populations, distinguishing three main groups: a first group that included populations from northwestern Spain, a second group with populations from northeastern Spain and a third group that included populations from southeastern Spain. Likewise, weak genetic differentiation was detected among geographically proximate populations contained in Groups II and III. Furthermore, there were seven more populations, from central and southern Spain, that were not grouped according to geographical proximity. Thus, there were some geographically close populations that showed high genetic distances, such as SP04 and SP05 (0.137), both from southern Spain (Fig. 2). The differentiation detected between populations SP04 and SP11 (0.102) was high, considering that both populations are located in the same region and separated by only 16 km.

The results obtained with STRUCTURE software were congruent with the clustering obtained with principal coordinate analysis. The most probable division with the strongest support in terms of log-likelihood values was detected at $K$=3, whereas for $K$>3 the results were not consistent and the membership analyses were unstable among runs. For $K$=3, we detected three clusters that corresponded to the three main groups identified in principal coordinate analysis (Fig. 3). All southern populations were assigned to Cluster I, except SP04; Cluster II grouped populations from northeastern Spain (assigned to Group II in the analysis of principal coordinates) along with SP06 and SP15 from central Spain; and Cluster III, which included populations from northwestern Spain (belonging to Group I) and population SP04 from southern Spain. Furthermore, populations SP07 and SP14 displayed high degree of admixtures of the three clusters (Fig. 3).

Some geographical trends in genetic diversity were found for the assessed material (Table 4). The expected heterozygosity (He) was not significantly correlated with latitude, longitude and altitude, whereas Ar and $F_{IS}$ were significantly negatively correlated with latitude and longitude (r = -0.594 and r = -0.597, $P < 0.05$ respectively).

Results obtained by Monmonier's algorithm are shown in Fig. 4. Overall, these results were congruent with those obtained with the Bayesian and principal coordinate analyses. Along the western edge, from north to south, a large genetic surface was found indicating low discontinuities. In this respect, the only boundary detected was around SP04, which was very different from the other three populations situated in the southern region (Fig. 4a). Regions of elevated genetic differentiation were detected among northwestern, southwestern and the central part of the southern region populations, an area in which the species distribution was more continuous. Nevertheless, some discontinuities were found, mainly due to the large differentiation

showed by populations situated in the central part of the southern region (Fig. 4b). By contrast, populations situated in northeastern and southeastern showed the lowest level of differentiation among them (Fig. 2 and Fig. 4b).

## Discussion

Our results highlighted that Spanish chestnut populations are characterised by high genetic diversity, confirming the results obtained in previous studies conducted on European chestnut populations (Martín et al. 2010) and traditional Spanish chestnut varieties (Martín et al. 2009; Pereira-Lorenzo et al. 2010). We identified 92 alleles with high reliability, resolution and higher variation compared with other markers such as ISSR (Mattioni et al. 2008) and isozymes (Fernandez-Lopez and Monteagudo 2010).

According to Mantel test, no significant correlation between genetic differentiation and geographic distance was found in the assessed populations. However, PCA and STRUCTURE analyses detected a clear geographic pattern showing three different groups of populations in northwestern, northeastern and southeastern Spain. This could indicate that although there is a geographic structure in these populations, it is not necessarily strong, because the genetic diversity among sites did not follow a continuous pattern. Furthermore, the low genetic distance between populations contained in the three groups probably indicates that populations belonging to each of these groups share a common gene pool due to their geographic proximity. Conversely, populations situated in the central part of southern Spain were the most divergent, showing the highest genetic differentiation between them and with respect to the rest of the populations.

Northern and southern populations displayed different genetic composition, and some differentiation was also detected among western and eastern areas. Different alleles

were detected in northern and southern populations, including 10 private alleles in the northern populations and 13 in the southern populations. Part of the differences found between west and east could be due to the fact that distribution of the species in the western region was continuous, while in the eastern region chestnut was only located in northern and southern sides without intermediate stands.

Considering these results, our sampling populations permitted us to formulate a hypothesis regarding the pattern of genetic structure of the species in Spain, suggesting the influence of both historical climate changes and human activity. Palynological data indicated that chestnut was a characteristic component of Iberian oak forests during the Tertiary period (Garcia-Anton et al. 1990, Costa et al. 1998), which seems to confirm the role of the Iberian Peninsula as a refugium during the *Würm* glaciation (Gomez-Sanz et al. 2002). Our results confirm that northwestern populations (Group I) could have their origin in a possible refugium area located in southern Galicia, as described by other authors (Krebs et al. 2004; Fineschi et al. 2000; Mattioni et al. 2008). We also detected another different group of populations situated in the northeast (Group II). Our results agree with those from pollen records (Huntley and Birks 1983; Garcia-Anton et al. 1990), and suggest that these populations constitute a second refugium in northeastern Spain. Although our results on populations in northern Spain support the hypothesis of two different refugia, it is not possible to exclude an overlapping human intervention, at least at the local level.

Conversely, a possible explanation for the origin of the southern populations is even more complex. First, a group was identified that comprised the southeastern material (Group III) and another highly divergent group of populations. Given that there are no certain records of pollen in this area, the most plausible hypothesis is that man influenced the genetic structure of this material. The extensive movement of chestnut

genetic material across Europe in the past could explain this finding (Columela 1979; Adua 1999; Conedera et al. 2004), considering that Andalusia (southern Spain) was one of the regions most influenced by Romans in the Mediterranean basin. An example of this human influence was clearly detected in population SP04 which, despite being only 16 km apart from SP11, was genetically close to the northwestern populations. It is highly probable that this population was planted from the northwestern gene pool. In this respect, understanding of the comprehensive genetic relationships between wild and cultivated chestnuts is needed to clarify this complex situation, as well as deeper studies based on the human role in the alteration of the spatial composition of the species.

Some geographical trends in genetic diversity were found for the assessed populations; Ar and $F_{IS}$ were significantly and negatively correlated with latitude and longitude, respectively. However, given the different origin of the populations, these data should be interpreted cautiously.

Our results, based in genetic landscape shape analysis, agree with results obtained with the other genetic structure approaches. Low discontinuities were detected in the distribution area of the species, with only one boundary around population SP04. Based on this pattern, it is quite clear that the large genetic distances found in the southern region are not due to the presence of a strict geographical barrier, but to the fact that these populations contain divergent alleles (six private alleles considering only the southern populations). In this respect, the high genetic differentiation detected in population SP04 must be emphasized.

In conclusion, our results highlight that the genetic structure of the current chestnut stands in Spain is the result of a natural colonization history and human driven domestication. Moreover, this study provides valuable baseline data that should allow

more in-depth studies of landscape genetics associated with this species and that can contribute to its conservation and management.

**Considerations for conservation and management**

We propose the three detected groups, along with the set of populations located in the southern region, as possible conservation zones of the species in Spain. Among them, conservation priority should be given to those populations that display the higher number of private alleles and maintain enough genetic variability. These populations could be well-adapted to local conditions, reinforcing the importance of conserving locally adapted genotypes. Moreover, further ecological and environmental studies could help to select them as conservation units. We conclude that Spanish populations play an important role in the genetic diversity of the species and therefore must be considered for the preservation of its genetic resources.

**Acknowledgements**

This research was supported by grants AGL2009-07931 and AGL2010-15147 from the Spanish Ministry of Science and Innovation and the European Regional Development Fund (FEDER) from the European Union. The first author is grateful to the «Alfonso Martin Escudero» Foundation for a postdoctoral fellowship and the hosting Institute of the fellowship, Agroenvironmental and Forest Biology Institute from Italian National Research Council.

**References**

Adua M (1999) The sweet chestnut throughout history from the Miocene to the third millennium. Acta Hort 494: 29-36

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. Nat Rev Genet 11:697-709

Allué JL (1990) Atlas fitoclimático de España. Monografías INIA, nº69. MAPA-INIA, Madrid, pp 221

Booy G, Hendricks RJJ, Smulders MJM, van Groenendael JM, Vosman B (2000) Genetic diversity and the survival of populations. Plant Biol 2: 379-395

Buck EJ, Russell K, Hadonou M, James CJ, Blakesley D (2003) Isolation and characterization of polymorphic microsatellites in European chestnut (*Castanea sativa* Mill.). Mol Ecol Notes 3: 239-241

Columela LJM (1979) De Res Rustica. Siglo I d.C. Los doce libros de Agricultura. Editorial Iberia, Barcelona.

Conedera M, Krebs P, Tinner W, Pradella M, Torriani D (2004) The cultivation of *Castanea sativa* Miller in Europe, from its origin to its diffusion on a continental scale. Veg Hist Archaebot 13:161-179

Costa M, Morla C, Sainz H (1998) Los bosques ibéricos. Una interpretación paleobotánica. Editorial Planeta, Barcelona.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of cluster of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14: 2611-2620

Fady B, Conord C (2010) Macroecological patterns of species and genetic diversity in vascular plants of the Mediterranean basin. Diversity Distrib 16: 53-64

Fernandez-Lopez J, Monteagudo AB (2010) Genetic structure of wild Spanish populations of *Castanea sativa* as revealed by isozyme analysis. Forest Systems 19: 156-169

Fineschi S, Taurchini D, Villani F, Vendarmin GG (2000) Chlororplast DNA polymorphism reveals little geograpgical structure in *Castanea sativa* Mill. (Fagaceae) throughout southern European countries. Mol Ecol 9: 1495-1503

Frankman R, Ralls K (1998) Conservation biology: inbreeding leads to extinction. Nature 392:441-442

Garcia-Anton M, Morla C, Sainz H (1990) Consideraciones sobre la presencia de algunos vegetales relictos terciarios durante el cuaternario en la Península Ibérica. Bol R Soc Esp Hist Nat (Sec. Biol.) 86:95-105

Gomez A, Vendramin GG, Gonzalez-Martinez S, Alia R (2005) Genetic diversity and differentiation of two Mediterranean pines (*Pinus halepensis* Mill. and *Pinus pinaster* Ait.) along a latitudinal cline using chloroplast microsatellite markers. Drivers Distribution 11:257-263

Gomez-Sanz V, Blanco-Andray A, Sánchez-Palomares O, Rubio-Sánchez A, Elena-Roselló R, Graña-Domínguez D (2002) Autoecología de los castañares andaluces. Inv Agrar Sist Rec F 11:205-226

Goudet J (2002) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3.2). Available from http://www.unil.ch/izea/softwares/fstat.html.

Gupta PK, Balyan IS, Sharma PC, Ramesh B (1996) Microsatellite in plants-a new class of molecular markers. Curr Sci 70:45-54

Hedrick PW (2005) A standardised genetic differentiation measure. Evolution 59:1633-1638

Holderegger R, Wagner HH (2008) Landscape Genetics. BioScience 58:199-207

Huntley B, Kirks HJB (1983) An atlas of past and present pollen maps for Europe: 0-13,000 years ago.Cambridge University Press, Cambridge.

Krebs P, Conedera M, Pradella M, Torriani D, Felber M, Tinner W (2004) Quaternary refugia of the sweet chestnut (*Castanea sativa* Miller): an extended palynological approach. Veg Hist Archaebot 13:145-160

Latta RG (2006) Integrating patterns across multiple genetic markers to infer spatial process. Landscape Ecol 21: 809-820

Lauteri M, Monteverdi MC, Sansotta A, Cherubini M, Spaccino L, Villani F. Küçük M (1998) Adaptation to drought in European chestnut. Evidences from a hybrid zone and from controlled crosses between drought and wet adpated populations. Acta Hort 494:345-353

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. Trends Ecol Evol 18:189-197

Manly BFJ (1985) The statistics of natural selection. Chapman and Hall, London.

Marinoni D, Akkak A, Bounous G, Edwards KJ, Botta R (2003) Development and characterization of microsatellite markers in *Castanea sativa* (Mill.). Mol Breeding 11:127-136

Martín MA, Alvarez JB, Mattioni C, Cherubini M, Villani F, Martín LM (2009) Identification and characterisation of traditional chestnut varieties of southern Spain using morphological and simple sequence repeats SSR markers. Ann. Appl. Biol. 154:389-398

Martín MA, Mattioni C, Cherubini M, Taurchini D, Villani F (2010) Neutral and adaptive genetic diversity in European chestnut populations by means of genomic and genic microsatellite markers. Tree Genet Genom 6:735-744

Mattioni C, Cherubini M, Micheli E, Villani F, Bucci G (2008) Role of domestication in shaping *Castanea sativa* genetic variation in Europe. Tree Genet Genomes 4:563-574

Miller MP (2005) Alleles in the Space (AIS): Computer software for the joint analysis of interindividual spatial and genetic information. J. Heredity 96:722-724

Monmonier M (1973) Maximun-differences barriers: an alternative numerical regionalization method. Geogr Anal 5:245-261

Namkoong G (2001) Forest Genetics: pattern and complexity. Can J Forest Res 31: 623-632

Nei M (1972) Genetic distance between populations. Amer Nat 106: 283-392

Oddou-Muratorio S, Demuse-Munsh B, Pélissier R, Gouyon PH (2004) Impacts of gene flow and logging history on the local genetic structure of scattered tree species, *Sorbus torminalis* L. Crantz. Mol Ecol 13: 3689-3702

Pautasso M (2009) Geographical genetics and the conservation of forest trees. Perspect Plant Ecol 11: 157-189

Pereira-Lorenzo S, Costa R, Ramos-Cabrer A, Ribeiro C, da Silva M, Manzano G, Barreneche T. 2010. Variation in grafted european chestnut and hybrids microsatellite reveals two main origins in the Iberian Peninsula. Tree Genet Genom 5: 701-715

Petit RJ, Latouche-Hallé C, Pemonge MH, Kremer A (2002) Chloroplast DNA variation of oaks in France and the influence of forest fragmentation on genetic diversity. Forest Ecol Manag 156: 115-129

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215-222

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945-959

Rivas-Martinez S (1987) Memoria del mapa de vegetación de España. 1:400.000. ICONA, MAPA, Madrid. pp 110

Saccheri I, Kuussaari M, Kankare M, Vikman P, Fortelius W, Hanski I (1998) Inbreeding and extintion in a butterfly metapopulation. Nature 392: 491-494

Schneider S, Roessli D, Excoffier L (2000) Arlequin: a software for population genetics data analysis, version 3.1. genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland

Shachak M, Boeken B, Groner E, Kadmon R, Lubin Y, Meron E, Neeman G, Perevolotsky A, Shkedy Y, Ungar ED (2008) Woody species as landscape modulators and their effect on biodiversity patterns. Bioscience 58: 209–221

Sork VL, Smouse PE (2006) Genetic analysis of landscape connectivity in tree populations. Landscape Ecol 21: 821-836

Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, Dezzani R, Delmelle E, Vierling L, Waits LP (2007) Putting the 'landscape' in landscape genetics. Heredity 98: 128-142

Tautz D, Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res 12:4127-4138

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535-538

Vendramin GG, Scotti I, Ziegenhagen B (2004) Microsatellites in Forest Tree Species: characteristics, identification and application. In: Molecular Genetics and Breeding of Forest Trees. Kumar S, Fladung M (eds). The Haworth Press, Inc. pp 429

Villani F, Sansota A, Cherubini M, Cesaroni D, Sbordoni V (1999) Genetic structure of natural populations of *Castanea sativa* in Turkey: evidence of a hybrid zone. J Evolution Biol 12:233-244

Watson DF (1992) Contouring: a guide to the analysis and display of spatial data. Pergamon Press, New York

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of populations structure. Evolution 38:1358-1370

Wright JP, Jones CG (2006) The concept of organisms as ecosystem engineers ten years on: progress, limitations, and challenges. BioScience 56 203–209

Yeh FC, Yang RC, Boyle TBJ, Ye ZH, Mao JX (1997). Popgene ver 1.32. The user-friendly software for population genetic analysis. Molecular Biology and Biotechnology Center, University of Alberta, Canada

1

Table 1. Description of each sampling population despicted in Figure 1.

| CODE | Population | Province | Location | Nº Trees | Latitude (N) | Longitude (W) | Altitude (m s.l.m.) |
|---|---|---|---|---|---|---|---|
| SP01 | Bubión | Granada | S | 14 | 36º56'46" | 3º21'08" | 1400 |
| SP02 | Castanyet | Girona | NE | 15 | 41º53'24" | 2º37'48" | 275 |
| SP03 | Costa Atlántica | Galicia | NW | 17 | 42º36'29" | 7º51'59" | 100 |
| SP04 | Gaucín | Málaga | S | 17 | 36º32'20" | 5º18'34" | 730 |
| SP05 | Güejar Sierra | Granada | S | 14 | 37º08'58" | 3º25'50" | 1144 |
| SP06 | Hervas | Cáceres | C | 14 | 40º15'36" | 5º12'01" | 967 |
| SP07 | Médulas | León | C | 15 | 42º27'00" | 6º03'00" | 975 |
| SP08 | Mieres | Asturias | NW | 15 | 43º13'01" | 5º45'50" | 450 |
| SP09 | Paterna | Almería | S | 14 | 37º01'26" | 2º56'34" | 1160 |
| SP10 | Prades | Tarragona | NE | 15 | 41º20'24" | 0º59'24" | 1050 |
| SP11 | Pujerra | Málaga | S | 12 | 36º35'38" | 5º8'37" | 870 |
| SP12 | Santa Elena | Jaen | S | 11 | 38º25'4" | 3º33'37" | 800 |
| SP13 | Sierra Faro | Galicia | NW | 21 | 43º17'09" | 8º22'11" | 660 |
| SP14 | Sierra Norte | Sevilla | S | 15 | 37º54'55" | 5º37'45" | 680 |
| SP15 | Valverde | Badajoz | C | 15 | 40º13'48" | 6º07'48" | 616 |
| SP16 | Viladrau | Girona | NE | 15 | 41º50'48" | 2º24' 00" | 900 |

S, south; NE, northeastern; NW, northwestern; C, centre.

2

3

4

1

Table 2 Description of the diversity for each of seven microsatellite loci.

| Locus | Size range | A | Ne | Ho | He | $F_{IS}$ |
|---|---|---|---|---|---|---|
| *CsCAT1* | 174-221 | 10 | 2.75 | 0.590 | 0.646 | -0.055 |
| *CsCAT2* | 195-247 | 17 | 8.50 | 0.775 | 0.882 | -0.104 |
| *CsCAT3* | 189-264 | 23 | 7.23 | 0.680 | 0.862 | 0.040 |
| *CsCAT6* | 158-196 | 14 | 9.98 | 0.736 | 0.897 | -0.014 |
| *CsCAT16* | 124-145 | 7 | 3.75 | 0.703 | 0.732 | -0.115 |
| *EMCs25* | 140-162 | 6 | 3.56 | 0.321 | 0.713 | 0.431** |
| *EMCs38* | 229-270 | 15 | 9.71 | 0.786 | 0.898 | -0.003 |
| Mean | | 13.14 | 6.50 | 0.656 | 0.804 | 0.026 |

A, number of detected alleles; Ne, effective number of alleles; Ho, observed heterozygosity; He, expected heterozygosity; $F_{IS}$, fixation index. Significant ** $P< 0.01$.

2

3

1

Table 3. Genetic diversity of the 16 chestnut populations investigated in this study.

| Pop | Nº Trees | Na | Ar | Ho | He | $F_{IS}$ | $F_{IS\ (6\ loci)}$ | $F_{ST}$ | $F_{ST'}$ | $F_{ST\ (6\ loci)}$ | $F_{ST'\ (6\ loci)}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| SP01 | 14 | 4.14 | 3.99 | 0.561 | 0.629 | 0.109 | 0.014 | | | | |
| SP02 | 15 | 4.86 | 4.70 | 0.819 | 0.722 | -0.145 | -0.144 | | | | |
| SP03 | 17 | 7.00 | 5.82 | 0.700 | 0.681 | -0.040 | -0.098 | | | | |
| SP04 | 17 | 5.57 | 4.87 | 0.487 | 0.680 | 0.280* | 0.180** | | | | |
| SP05 | 14 | 4.42 | 4.15 | 0.592 | 0.669 | 0.119 | 0.070 | | | | |
| SP06 | 14 | 6.00 | 5.58 | 0.633 | 0.703 | 0.107 | 0.013 | | | | |
| SP07 | 15 | 6.43 | 5.82 | 0.742 | 0.721 | -0.020 | -0.031 | | | | |
| SP08 | 15 | 4.86 | 4.56 | 0.733 | 0.713 | -0.040 | -0.063 | | | | |
| SP09 | 14 | 4.43 | 4.29 | 0.449 | 0.680 | 0.332* | 0.289** | | | | |
| SP10 | 15 | 4.14 | 3.97 | 0.790 | 0.667 | -0.210 | -0.232 | | | | |
| SP11 | 12 | 4.86 | 4.77 | 0.702 | 0.723 | 0.030 | 0.017 | | | | |
| SP12 | 11 | 5.43 | 5.43 | 0.636 | 0.695 | 0.084 | 0.098 | | | | |
| SP13 | 21 | 6.14 | 5.00 | 0.597 | 0.688 | 0.170 | 0.053 | | | | |
| SP14 | 15 | 5.71 | 5.51 | 0.676 | 0.750 | 0.090 | 0.051 | | | | |
| SP15 | 15 | 5.43 | 4.96 | 0.686 | 0.676 | -0.020 | -0.086 | | | | |
| SP16 | 15 | 5.14 | 4.73 | 0.657 | 0.662 | 0.010 | -0.015 | | | | |
| | | | | | | | | 0.152 | 0.141 | 0.145 | 0.137 |

Na, number of detected alleles; Ar, allelic richness; Ho, observed heterozygosity; He, expected heterozygosity; FIS, fixation index, $F_{IS\ (6\ loci)}$ fixation index of the six loci not affected by null alleles.**P< 0.01; $F_{ST}$, coefficient of differentiation among populations; $F_{ST'}$, standarised coeffcient of differentiation according Hedrick (2005); $F_{ST\ (6\ loci)}$, coefficient of differentiation among populations of the six loci not affected by null alleles; $F_{ST'\ (6\ loci)}$, standarised coeffcient of differentiation of the six loci not affected by null alleles.

2

3

1

Table 4 Matrix of correlation between geographical variables and diversity parameters.

| | Latitude | Longitude | Altitude |
|---|---|---|---|
| Ar | -0.594* | 0.289 | -0.491 |
| He | -0.301 | 0.133 | -0.453 |
| $F_{IS}$ | 0.103 | -0.597* | 0.366 |

Ar, allelic richness; He, expected heterozygosity; $F_{IS}$, fixation index. Significant * $P < 0.05$.

Suplementary Table. Pairwise Fst estimates among all populations of chestnut included in this study. Significant pairwise comparisons ($p < 0.05$) are included in bold.

| | SP01 | SP02 | SP03 | SP04 | SP05 | SP06 | SP07 | SP08 | SP09 | SP10 | SP11 | SP12 | SP13 | SP14 | SP15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SP01 | | | | | | | | | | | | | | | |
| SP02 | **0.109** | | | | | | | | | | | | | | |
| SP03 | **0.140** | **0.092** | | | | | | | | | | | | | |
| SP04 | **0.137** | **0.106** | **0.078** | | | | | | | | | | | | |
| SP05 | **0.049** | **0.115** | **0.139** | **0.120** | | | | | | | | | | | |
| SP06 | **0.125** | **0.054** | **0.081** | **0.096** | **0.119** | | | | | | | | | | |
| SP07 | **0.090** | **0.069** | **0.071** | **0.084** | **0.089** | **0.049** | | | | | | | | | |
| SP08 | **0.133** | **0.085** | **0.053** | **0.080** | **0.137** | **0.076** | **0.057** | | | | | | | | |
| SP09 | **0.065** | **0.113** | **0.150** | **0.123** | 0.021 | **0.130** | **0.103** | **0.143** | | | | | | | |
| SP10 | **0.136** | **0.044** | **0.104** | **0.120** | **0.126** | **0.080** | **0.082** | **0.107** | **0.133** | | | | | | |
| SP11 | **0.123** | **0.104** | **0.111** | **0.102** | **0.106** | **0.103** | **0.091** | **0.116** | **0.103** | **0.113** | | | | | |
| SP12 | **0.101** | **0.082** | **0.099** | **0.095** | **0.090** | **0.072** | **0.054** | **0.092** | **0.096** | **0.120** | **0.113** | | | | |
| SP13 | **0.140** | **0.116** | **0.030** | **0.076** | **0.143** | **0.088** | **0.072** | **0.039** | **0.158** | **0.124** | **0.125** | **0.115** | | | |
| SP14 | **0.116** | **0.075** | **0.087** | **0.076** | **0.101** | **0.058** | **0.060** | **0.088** | **0.103** | **0.094** | **0.063** | **0.080** | **0.101** | | |
| SP15 | **0.110** | **0.067** | **0.107** | **0.110** | **0.122** | **0.037** | **0.072** | **0.105** | **0.129** | **0.091** | **0.116** | **0.077** | **0.116** | **0.078** | |
| SP16 | **0.126** | 0.026 | **0.122** | **0.124** | **0.137** | **0.071** | **0.081** | **0.108** | **0.129** | **0.080** | **0.142** | **0.075** | **0.140** | **0.097** | **0.077** |

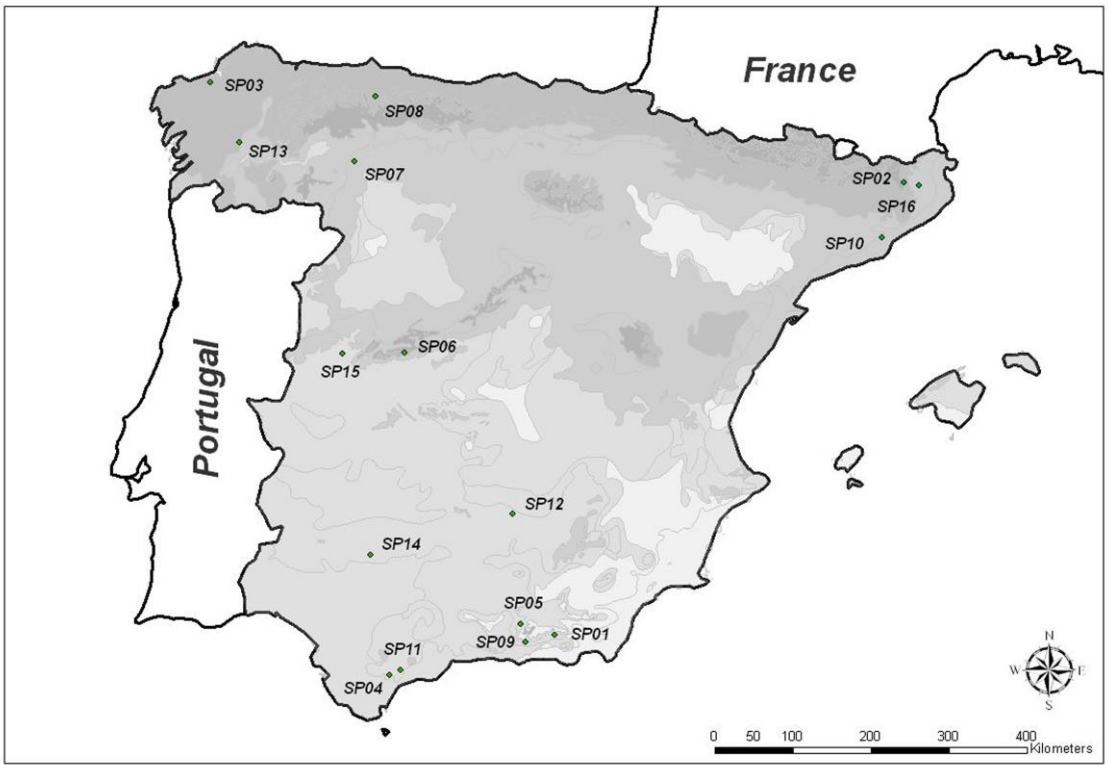
**Figure Caption**

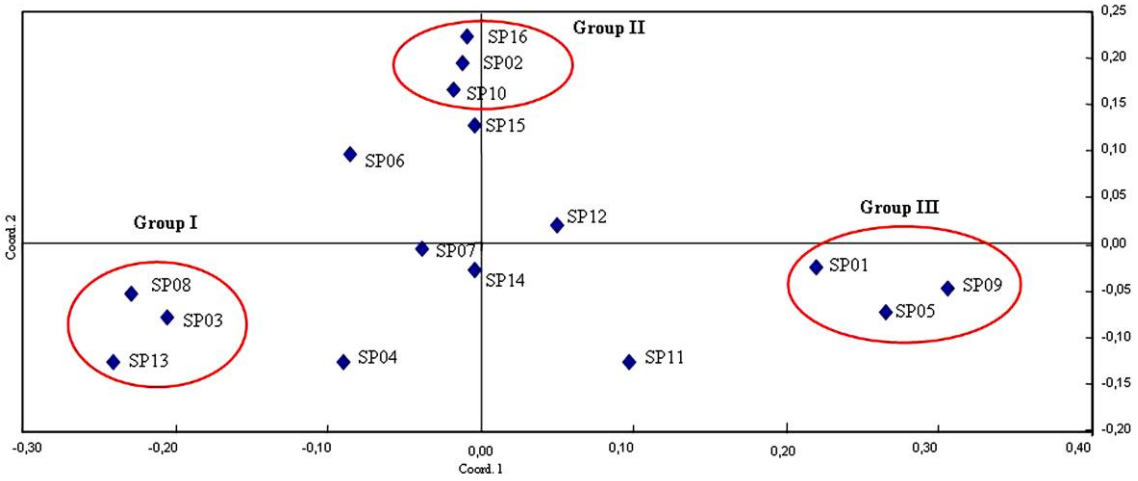
**Fig. 1** Map illustrating the location of the 16 Spanish populations of *Castanea sativa* Mill. examined in the study.

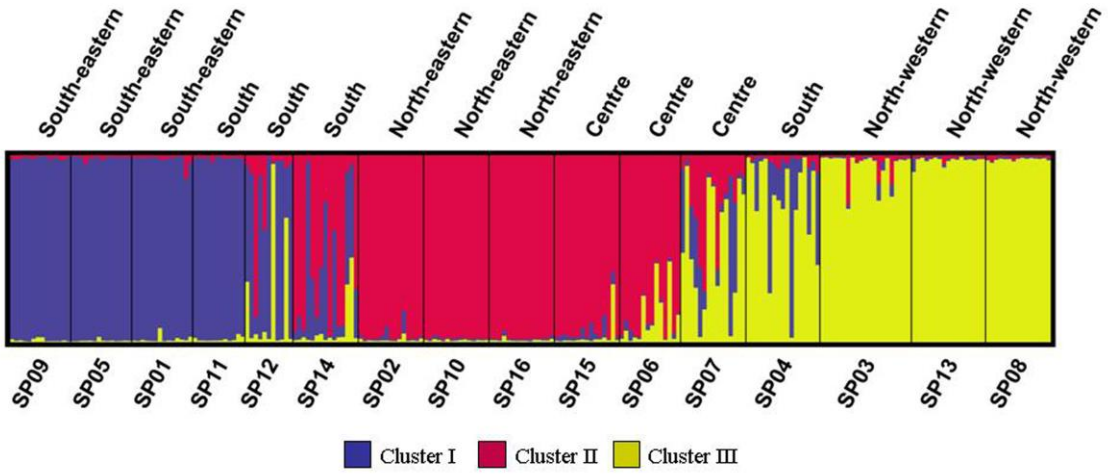
**Fig. 2** Principal coordinate analysis of the 16 chestnut populations from Spain. Group I, populations from northwestern Spain; Group II, populations from centre and northeastern Spain; and Group III, populations from southeastern Spain.

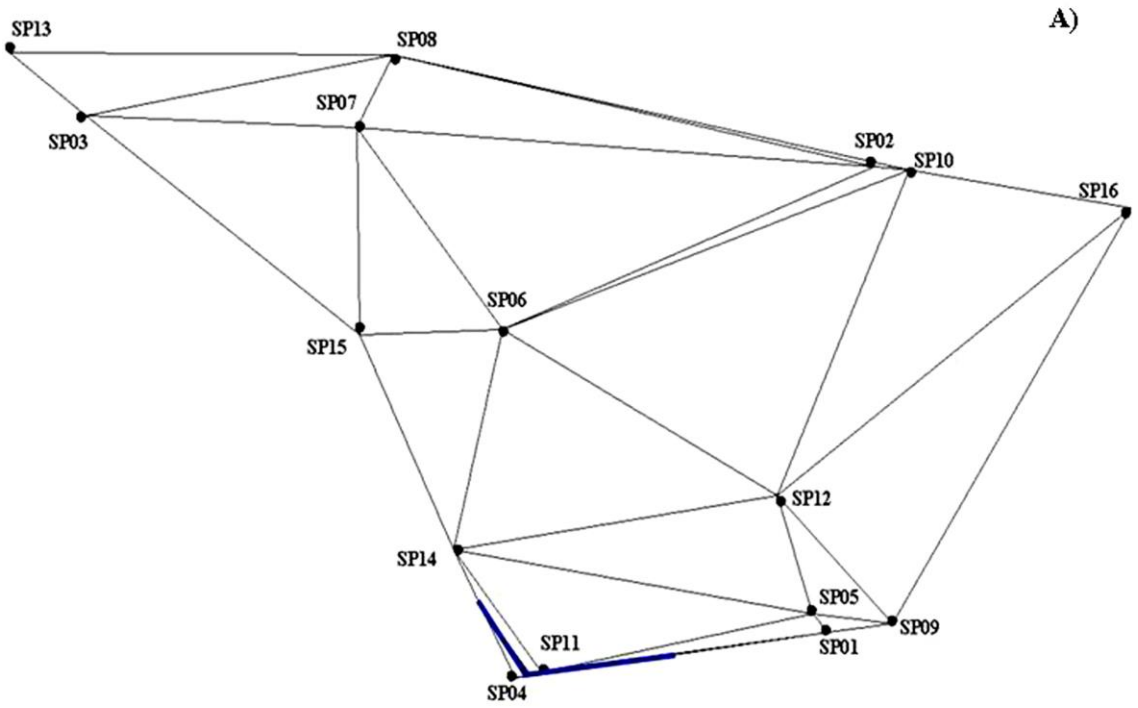
**Fig. 3** Genetic relationship among the 16 populations estimated using STRUCTURE (Pritchard et al 2000) and the data of the seven SSR loci for $K$=3. The different populations are separated by a vertical black line and are labelled according to Table 1.

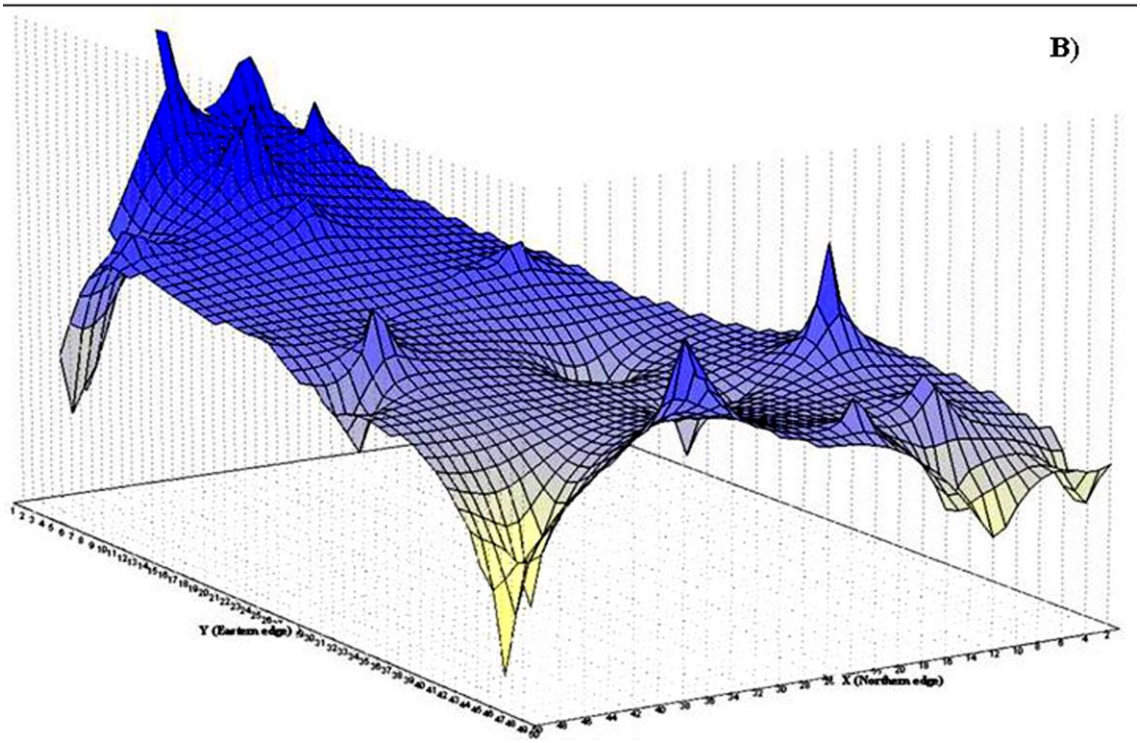
**Fig. 4** a) Diagrammatic representation of the Delaunay triangulation-based connectivity network used for Genetic Landscape Shape interpolation, showing inferred barriers to gene flow from the Monmonier´s analysis. The vertices of triangles on the connectivity network represent sampling populations assessed. b) Results of Genetic Landscape Shape interpolation analysis using a 50 × 50 grid. X and Y axes correspond to geographic locations within the populations analysed in the study. Z axe reflects the genetic distances.


France
Portugal
SP03
SP08
SP13
SP07
SP02
SP16
SP10
SP06
SP15
SP12
SP14
SP05
SP09
SP01
SP11
SP04
N
W
E
S
0 50 100 200 300 400
Kilometers


Group II
SP16
SP02
SP10
SP15
SP06
Group I
SP12
Group III
SP07
SP14
SP01
SP09
SP08
SP03
SP05
SP13
SP04
SP11
Coord. 2
Coord. 1
0,25
0,20
0,15
0,10
0,05
0,00
-0,05
-0,10
-0,15
-0,20
-0,30
-0,20
-0,10
0,00
0,10
0,20
0,30
0,40


South-eastern
South-eastern
South-eastern
South
South
South
North-eastern
North-eastern
North-eastern
Centre
Centre
Centre
South
North-western
North-western
North-western
SP09
SP05
SP01
SP11
SP12
SP14
SP02
SP10
SP16
SP15
SP06
SP07
SP04
SP03
SP13
SP08
Cluster I
Cluster II
Cluster III


A)
SP13
SP08
SP07
SP03
SP02
SP10
SP16
SP06
SP15
SP12
SP14
SP05
SP09
SP11
SP01
SP04


B)
Y (Eastern edge)
X (Northern edge)