		M	E	C	3	2	7	5	**Operator:** Zheng Yuexia	**Dispatch:** 01.02.07	**PE:** Ann Cowie
	Journal Name				Manuscript No.				**Proofreader:** Chen Xiaoming	**No. of Pages:** 16	**Copy-editor:** Lulu Stader

Molecular Ecology (2007) doi: 10.1111/j.1365-294X.2007.03275.x

Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers

GABRIELE BUCCI,* SANTIAGO C. GONZÁLEZ-MARTÍNEZ,† GRÉGOIRE LE PROVOST,‡ CHRISTOPHE PLOMION,‡ MARIA MARGARIDA RIBEIRO,§ FEDERICO SEBASTIANI,¶ RICARDO ALÍA† and GIOVANNI GIUSEPPE VENDRAMIN*

**Istituto di Genetica Vegetale, Sezione di Firenze, Consiglio Nazionale delle Ricerche, via Madonna del Piano 10, 50019 Sesto Fiorentino (FI), Italy, †Departamento de Sistemas y Recursos Forestales, CIFOR — INIA, Carretera de La Coruña km 7.5, 28040 Madrid, Spain, ‡INRA, UMR BIOGECO, 69 route d'Arcachon, 33610 Cestas, France, §Escola Superior Agrária, Unidade Departamental de Silvicultura e Recursos Naturais, 6001-909 Castelo Branco, Portugal, ¶Dipartimento di Biotecnologie Agrarie, Genexpress, Università di Firenze, Via della Lastruccia 14, 50019 Sesto Fiorentino (FI), Italy*

Abstract

Some 1339 trees from 48 *Pinus pinaster* stands were characterized by five chloroplast microsatellites, detecting a total of 103 distinct haplotypes. Frequencies for the 16 most abundant haplotypes ($p_k > 0.01$) were spatially interpolated over a lattice made by 430 grid points. Fitting of spatially interpolated values on raw haplotype frequencies at the same geographical location was tested by regression analysis. A range-wide 'diversity map' based on interpolated haplotype frequencies allowed the identification of one 'hotspot' of diversity in central and southeastern Spain, and two areas of low haplotypic diversity located in the western Iberian peninsula and Morocco. Principal component analysis (PCA) carried out on haplotypes frequency surfaces allowed the construction of a colour-based 'synthetic' map of the first three PC components, enabling the detection of the main range-scale genetic trends and the identification of three main 'gene pools' for the species: (i) a 'southeastern' gene pool, including southeastern France, Italy, Corsica, Sardinia, Pantelleria and northern Africa; (ii) an 'Atlantic' gene pool, including all the western areas of the Iberian peninsula; and (iii) a 'central' gene pool, located in southeastern Spain. Multivariate and amova analyses carried out on interpolated grid point frequency values revealed the existence of eight major clusters ('gene zones'), whose genetic relationships were related with the history of the species. In addition, demographic models showed more ancient expansions in the eastern and southern ranges of maritime pine probably associated to early postglacial recolonization. The delineation of the gene zones provides a baseline for designing conservation areas in this key Mediterranean pine.

Keywords: cpSSR, gene zones, geostatistics, haplotype diversity map, maritime pine

Received 3 July 2006; revision received 21 November 2006; accepted 8 January 2006

Introduction

Maritime pine (*Pinus pinaster* Ait.) is an economically important species for timber and pulp production in southwestern Europe (Brendel *et al*. 2002; Pot *et al*. 2005). It is also used for afforestation in several European countries, with extensive seed transfer among breeding zones (Baradat & Marpeau-Bezard 1988; Devy-Vareta 1988) and outside its native range (e.g. Australia, New Zealand, South Africa). Recently, the pest *Matsucoccus feytaudi* has been spreading into the eastern part of maritime pine range, provoking population decline in northern Italy (Covassi & Binazzi 1992) and Corsica (Burban *et al*. 1999). The species' Mediterranean populations are also threatened by recurrent wildfires (often associated with replacement by other forest trees, such as Aleppo pine and Eucalyptus), generally from human origin and triggered by global warming; and, in some cases such as Morocco, by overexploitation and absence of natural regeneration (Wahid *et al*. 2004 and references therein). The above issues make *P. pinaster* a model species for population genetic studies aimed at the

Correspondence: G. G. Vendramin, Fax: +39 055 522 5729; E-mail: giovanni.vendramin@igv.cnr.it

conservation of the species' evolutionary/adaptive potential and the exploitation of its breeding resources.

In the last decades, a fairly large amount of genetic and phylogeographical information on the species has been reported in the literature on several fields of investigation, namely water-stress resistance and ecophysiology (Nguyen & Lamant 1989; Loustau *et al*. 1995), growth-related traits and quantitative genetics (Alía *et al*. 1997; Chambel 2006), population genetics and phylogeography (Vendramin *et al*. 1998; González-Martínez *et al*. 2001; Derory *et al*. 2002; Burban & Petit 2003; Gómez *et al*. 2005), palinology and palaeoecology (Carrión & Van Geel 1999; Carrión *et al*. 2003), pest and disease resistance (Desprez & Baradat 1991) and patterns of DNA sequence polymorphism (Pot *et al*. 2005), including review papers (González-Martínez *et al*. 2004; Ribeiro *et al*. 2006).

The presence of a *P. pinaster* centre of origin in the southwest of the Iberian peninsula at the end of the Pliocene (3 Ma) was hypothesized by Baradat & Marpeau-Bezard (1988) and supported by fossil findings (Teixeira 1945). The authors drew a different picture for the migration pathways of the species before and after the last glaciation. The preglacial hypothesis supposes the presence of three distinct pathways towards the north of Portugal, Spain and France, towards the south of Spain, France and Italy, and towards the north of Africa. The postglacial hypothesis assumes that migration occurs only along the first two pathways. Moreover, the authors claim that the successive ice ages during the Pleistocene had stopped the northerly advance of *P. pinaster* several times, and eventually reduced its presence to scattered refugia in the south of the Iberian peninsula, particularly during the Pleniglacial (0.35 Ma). More recently, Burban & Petit (2003) reported three broad regions fixed for different (maternally inherited) mitotypes: an eastern maternal lineage (Catalonia, southeastern France, Corsica, Italy, Pantelleria, Tunisia and Algeria), a western one (most of the Iberian peninsula, continental France and Punta Cires in northern Morocco) and a Moroccan specific maternal lineage. Other genetic markers [terpenes, isozymes, nuclear simple sequence repeats (SSRs) and amplified fragment length polymorphisms (AFLPs)] have shown a finer structure of genetic variation in maritime pine (see reviews in González-Martínez *et al*. 2004; Ribeiro *et al*. 2006).

Knowledge of geographical distribution of genetic resources is a basic step for the implementation of any conservation or breeding activity. Therefore, the detection of intraspecific biodiversity reservoirs at the range level may lead to the identification of gene reserve and breeding source populations (Bucci & Vendramin 2000). Furthermore, the range-scale delineation of genetically homogeneous regions (henceforth: 'gene zones') and the detection of the main genetic discontinuities within the species' range may provide a genetic baseline to be combined with climatic, edaphic and management information for the definition of breeding zones for the species (Westfall & Conkle 1992; Yang *et al*. 1998; Bucci & Vendramin 2000; Hamann *et al*. 2000).

Whether the recognition of distinct gene pools or provenances using neutral molecular markers has any meaning or not in the identification of adaptive/evolutionary potential sources for the species is still under debate by the scientific community (e.g. Taylor & Dizon 1999; see also Green 2005 for a review on the concept of evolutionary significant units — ESUs *sensu* Moritz 1994 — and Newton *et al*. 1999 for the application of this concept to forest trees). However, recent studies reveal consistency between molecular marker-based gene zones and classification of Norway spruce provenances based on growth traits (Bucci & Vendramin 2000), confirming previous studies by Giertych (1977, cited in Vidakovic 1991) and by Lagerkrantz & Ryman (1990). Adaptive trait-linked markers in *Picea abies* have been reported to show frequencies significantly different among gene pools previously recognized by molecular markers (Bozhko *et al*. 2003). Furthermore, a recent study on the same species based on a set of genome-wide mapped molecular markers revealed a different genome organization and hints at differential local adaptation among the above geographical domains (Acheré *et al*. 2005). However, only weak correspondence has been found in maritime pine between broad gene zones identified by different maternal lineages [mitochondrial DNA (mtDNA)] and a number of growth traits (González-Martínez *et al*. 2004) and, at least within the Spanish range (see González-Martínez *et al*. 2005), variation in quantitative traits was not correlated with that found using genetic markers.

A number of geostatistical techniques have been devised and applied to a wide range of fields (Piazza *et al*. 1981; Barbujani 1988; Guenni & Hutchinson 1998; Higdon 1998; Nalder & Wein 1998; Monastiez *et al*. 1999; Nanos *et al*. 2004, 2005). Geostatistical methods have been applied to population genetic analysis of forest tree species with the aim of predicting the behaviour of genetic parameters at unsampled points based on parameter values at neighbouring sampling locations and to study clinal patterns of variation (Westfall & Conkle 1992; Petit *et al*. 1997; Gömöry *et al*. 1998; Le Corre *et al*. 1998; Bucci & Vendramin 2000; González-Martínez *et al*. 2001). Gene frequencies are often affected by a fairly high sampling error, especially if they are based on a small number of individuals (Cavalli-Sforza *et al*. 1994). Moreover, unequal geographical distribution of sampling sites and the existence of geographical barriers may hamper the detection of the main trends in genetic variation across a species' natural range. The use of spatial interpolation methods helps detect the main geographical genetic trends by smoothing local erratic variation at sampled points. Local stochastic variation can result from sampling errors, planting of nonlocal material or long-range pollen movement (Bucci & Vendramin 2000).

In this work, we report the results of the analysis of 48 stands of *Pinus pinaster* spread throughout its current

Table 1 Sixteen most common haplotypes detected. Haplotype composition refers to chloroplast microsatellites Pt71936, Pt30204, Pt87268, Pt15169 and Pt36480, respectively (Vendramin *et al*. 1996). Haplotypes showing overall frequency (p_k) lower than 0.01 are not reported individually

Label	Haplotype	Counts	Frequency
H01	143/144/163/114/145	270	0.202
H02	143/144/163/113/145	264	0.197
H03	143/143/163/114/145	86	0.064
H04	143/145/163/114/145	78	0.058
H05	143/144/163/115/145	74	0.055
H06	143/146/163/114/145	47	0.035
H07	143/144/162/114/145	46	0.034
H08	143/144/163/113/146	42	0.031
H09	143/144/162/115/145	29	0.022
H10	143/145/163/114/140	27	0.020
H11	143/144/164/116/145	25	0.019
H12	143/143/163/113/145	22	0.016
H13	143/143/162/114/145	20	0.015
H14	143/143/163/115/145	20	0.015
H15	143/145/163/115/145	19	0.014
H16	144/144/163/114/145	17	0.013
rare haplotypes ($p_k < 0.01$)		253	0.190

range by the use of five polymorphic paternally inherited chloroplast microsatellite markers [chloroplast simple sequence repeats (cpSSRs)]. Spatial interpolation methods were applied to the 16 most common haplotypes in order to obtain haplotype frequency surfaces based on 430 grid points regularly distributed over the current range of the species. Multivariate analysis was applied on spatially interpolated haplotype frequencies to detect the main range-wide trends of genetic variation and the identification of significant discontinuities within the *P. pinaster* range of distribution. The main goals of this investigation were: (i) to describe the distribution of genetic diversity and the geographical structure of variation at the range-wide level of a key Mediterranean pine species; (ii) to identify populations or regions likely to share the same evolutionary history in terms of origin from common glacial refugia and colonization along migration routes ('gene pools') and their contact zones; and (iii) to provide a baseline to define and characterize genetically homogeneous zones (i.e. gene zones) within the *P. pinaster* current range that may be used with range-wide information on ecology, growth traits, management etc., for the definition of conservation and breeding units.

Materials and methods

Plant material, DNA extraction and scoring of polymerase chain reaction (PCR) products

Some 1339 individuals from 48 *Pinus pinaster* populations (mean: 27.89 ± 1.95 individuals per population) sampled across the species' full natural range were analysed (Table S1, supplementary material). Total DNA was extracted from needles according to Doyle & Doyle (1990) or by a modification of the procedure described by Ziegenhagen *et al*. (1995), using the QIAamp Blood Kit (QIAGEN). Five chloroplast microsatellite markers (Pt71936, Pt30204, Pt87268, Pt15169 and Pt36480; Vendramin *et al*. 1996) were selected because of high polymorphism and product-size differences, the latter allowing multiplex by size. Polymerase chain reaction (PCR)-amplification was performed using a Perkin Elmer 9600 thermal cycler following the procedures reported in Vendramin *et al*. (1996). Amplification reactions were automatically prepared using a Biomek 2000 robotic workstation. PCR products were mixed with internal size standards (50, 150 and 200 bp) and separated using 6% denaturating polyacrilamide gels on an ALF automatic sequencer (Amersham). Automatic sizing of the amplified fragments was performed using Fragment Manager software version 1.2 (Amersham).

Haplotype diversity

Haplotypes were inferred from the individual allele size profiles at the five cpSSR fragments analysed. The 103 detected haplotype were sorted by their overall frequency (p_k), and the 16 most common haplotypes were selected for further analyses (Table 1). Rare haplotype frequencies ($p_k < 0.01$) were pooled together and considered as a single haplotype frequency.

The effective number of haplotypes for each population was calculated as $n_e = (n-1)/(n\sum p_k^2 - 1)$ (Pamilo 1993; Nielsen *et al*. 2003), while the unbiased haplotype diversity was estimated as $H_e = [n/(n-1)](1 - \sum p_k^2)$ (Nei 1987), where p_k is

the frequency of the *k*-th haplotype and *n* is the number of individuals analysed. Within-population haplotypic diversity was also estimated calculating the parameter S_w (Slatkin 1995).

Pairwise haplotypic difference between individuals belonging to the same population (x_{ij}) was calculated as the number of mutational steps intervening between their haplotypes (Slatkin 1995):

$$x_{ij(i<j)} = \sum_{k=1}^{L} |a_{ik} - a_{jk}|,$$

where a_{ik} and a_{jk} are the variants' lengths at the *k*-th cpSSR region carried by the *i*-th and *j*-th individuals, and *L* is the total number of cpSSR regions compared. The number of mutations between any two haplotypes is a random variable drawn from a Poisson distribution (Slatkin & Hudson 1991; Slatkin 1995). Goodness-of-fit of within-population x_{ij} distribution with Poisson expectations was tested by a Kolmogorov–Smirnov test ($\alpha = 0.05$). The frequency distribution of the observed within-population pairwise x_{ij} values was calculated independently for each population.

Spatial interpolation of haplotype frequencies

Haplotype frequencies (p_k) were transformed by arcsinv p_k and standardized with mean = 0 and SD = 1. A different standardization method was also applied (Cavalli-Sforza *et al*. 1994) but it did not provide any remarkable difference and therefore was not used for further data analysis.

Grid creation and interpolation method. Spatial interpolation of haplotype frequency values was carried out on a regular square lattice of 430 grid points, with interleaving distance of 0.3×0.3 degrees longitude by latitude (Fig. 1). The lattice covers the whole natural range of *P. pinaster*, with a surrounding buffer area of 50 km.

For each haplotype, grid points frequency values were calculated as follows (Shepard 1968; Cavalli-Sforza *et al*. 1994):

$$\hat{p}_{jk} = \left(\sum_{i=1}^{N} 1/d_{ij}^2 \cdot p_{ik} \right) \Big/ \left(\sum_{i=1}^{N} 1/d_{ij}^2 \right)$$

where $\hat{p}_{jk}$ is the frequency estimate at the *j*-th grid point for the *k*-th haplotype, p_{ik} is the frequency of the *k*-th haplotype observed at the *i*-th sampling point, d_{ij} is the distance (in radians) of the *j*-th grid point from the *i*-th sampling point, and *N* is the total number of sampling points (stands) considered. Haplotype-frequency estimation at a given grid point started using the information of the nearest three sampling points, adding step-by-step the more remote ones, as far as the difference in estimates between subsequent steps was lower than a threshold (arbitrarily set at 0.025). The above method is isotropic (all directions in space are equally considered), locally adaptive (uses a number of neighbours proportional to the local heterogeneity of frequencies) and model-free (any theoretical formula of frequency variation in space is assumed), and it has proven to provide estimates in good agreement with those obtained with more sophisticated methods (e.g. *kriging*; see Cavalli-Sforza *et al*. 1994).

Validation of haplotype frequency surfaces. Goodness-of-fit of the interpolated haplotype frequency surfaces on the original (stand) sampling frequencies was verified as follows. For each sampling site and haplotype, the mean interpolated frequency value of the closest four grid points surrounding a given sampling point was considered as the 'expected' frequency of that haplotype at that geographical location. A linear regression analysis was carried out independently for each haplotype (PROC REG routine of the SAS/STAT package), using the above 'expected' values as predictors of the original observed frequencies (the observed haplotype frequency of the stand at the same location on the map). Mean absolute deviation of expected from observed values and the number of sampling frequencies falling outside the 95% confidence limits of the regression function were then recorded.

Principal component analysis (PCA) and identification of the main gene pools. Principal component analysis (PCA, PROC PRINCOMP of the SAS/STAT statistical package) was applied on the lattice dataset (430 records × 16 variables) after transformation and standardization as described above. A 'synthetic' map of the main gradients of genetic variation for *P. pinaster* across the range was obtained by combining together the information from the first three principal components (PCs) of each grid point using an RGB coding (red = PC1, green = PC2 and blue = PC3). Grid points PC scores (eigenvalues) were proportionally rescaled to a range from 0 (no colour) to 255 (full colour). The transparency of the whole PC layer was set as inversely proportional to the variance accounted for the corresponding PC axis. PC layers were then combined using a GIS software package (TNTmips©, Microimage LTD) and plotted on a map. Using this method, the relative intensity of the three main colours indicates approximately the position of a given grid point along the main gradients of haplotype frequency variation, as well as its distance to the PC extremes.

Finally, the 16 most common spatially interpolated haplotype frequencies analysed at the 430 grid points were used to assess genetic diversity (H_e) across the species full natural range. Expected heterozygosity values (H_e) were then standardized with mean = 0 and SD = 1 and plotted on a map using a GIS software (TNTmips©, Microimage LTD).

Identification and characterization of gene zones

Cluster analysis and gene-zone identification. Cluster analysis was carried out on the 430 grid points haplotype frequencies using the nonparametric density method implemented in the PROC MODECLUS routine of the SAS/STAT package.

Colour image

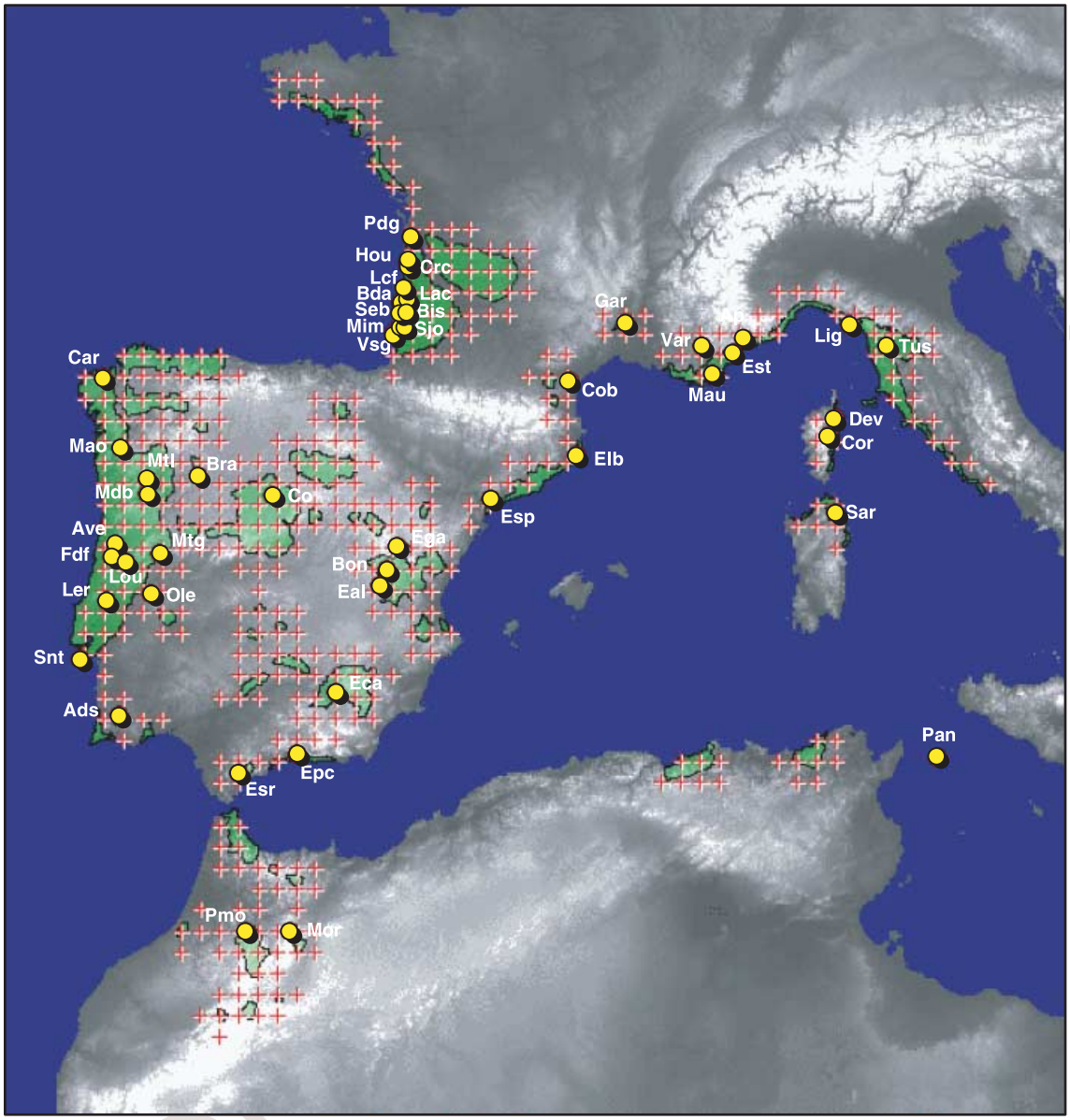

Fig. 1 Geographical distribution of the 48 stands sampled in this investigation (*yellow circles* on the map) and of the 430 grid points (0.3 × 0.3 degrees longitude by latitude — represented by *red crosses* on the map) overlying the current range of *P. pinaster* (sketched in *green* — from: Petit *et al*. 2003, redrawn) and used for spatial interpolation of haplotype frequencies. For more details, see *Materials* and *Methods*.

The number of clusters was determined by repeatedly joining the least significant cluster with neighbouring clusters until all remaining clusters were significant at $\alpha = 0.05$ (option TEST/JOIN of the PROC MODECLUS routine).

The grouping obtained was further tested by discriminant analysis (PROC DISCRIM routine) on the 16 interpolated haplotype frequencies. Grid points showing a posteriori probability of membership to the assigned cluster lower than 0.95 were considered as unclassified. A dendrogram of the above clusters was obtained using the PROC CLUSTER routine on a pairwise Mahalanobis distance matrix (as from discriminant analysis) and the Ward's minimum-variance clustering method. Robustness of the above grouping and genetic discontinuities among geographically adjacent clusters were tested by pairwise analysis of molecular variance (AMOVA) using the ARLEQUIN version 3.0 software package (Excoffier *et al*. 2005).

Sampled stands were assigned to one of the identified gene zones based on their geographical location, and their (observed) haplotype frequencies compared with mean haplotype frequencies of the parent cluster (expected). Goodness-of-fit of observed and expected haplotype frequency cumulative distributions was tested with the Kolmogorov–Smirnov nonparametric test. Stands showing

significant deviations from expectations based on spatially interpolated haplotype frequencies were considered outliers (i.e. not belonging to the local gene zone represented by the corresponding grid points cluster).

Characterization of gene zones. Differences in population gene diversity among gene zones (outliers excluded) were tested using a nonparametric Kruskal–Wallis test. Gene zones were further characterized by computing mismatch distributions using the ARLEQUIN version 3.0 software package (Excoffier *et al*. 2005). Mismatch distributions are frequently used to describe demographic processes such as population growth (Hudson & Slatkin 1991; Rogers & Harpending 1992) or range expansions (Ray *et al*. 2003; Excoffier 2004) and can provide valuable biogeographical information when applied to the comparison of gene pools associated with different glacial refugia or historical processes. However, only recently their applicability to chloroplast microsatellite data has been demonstrated, both by extensive simulation and empirical data (Navascués *et al*. 2006). First, chloroplast microsatellite haplotypes were binary coded: the number of repeats was coded with '1' and shorter alleles were coded by filling the difference in repeats with '0' (Navascués *et al*. 2006). Then, mismatch distributions were computed and the demographic model of Rogers & Harpending (1992) was fitted following Schneider & Excoffier (1999 — see also Excoffier *et al*. 2005). The goodness-of-fit of the demographic model was tested using 10 000 bootstraps and τ values for each gene pool ($\tau = 2\mu t$, where μ is the mutation rate and t is the number of generations since expansion) were computed.

Results

Haplotype diversity

Overall, 103 different haplotypes were detected in this investigation, with a mean of 10.60 ± 3.76 per stand (Table 2). Visual inspection of the single haplotype distribution revealed, in some cases, a strongly patchy distribution over the species range (Fig. S1, supplementary material). Out of the 103 haplotypes, 41 (39.8%) were specific for a single population (private haplotypes), and had very low frequency (mean: 0.0377 ± 0.0016; range: 0.0222–0.0555). Private haplotypes were found in 20 stands, mostly distributed in the east of Spain (six stands) and the southern and eastern part of the species range (Fig. S2, supplementary material). The unbiased haplotypic diversity was remarkably lower than the overall mean in the western part of the Iberian peninsula, and the highest values were found in southwestern Spain (Fig. 2A). Maximum within-population haplotypic variance (S_w) was found in stands from eastern and central Spain and from southwestern France (Landes region), while lowest values of S_w were observed for stands from

Table 2 Main within-population genetic parameters for the 48 stands analysed in this study. *N*, stand sample size; *H*, number of haplotypes; n_e, effective number of haplotypes (Nielsen *et al*. 2003); H_e, haplotypic diversity (Nei 1987); x_{ij}, mean number of pairwise differences among haplotypes (Slatkin 1995), significant deviations of x_{ij} from a Poisson distribution are indicated by *: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$

Stand	*N*	*H*	n_e	H_e	x_{ij}
Ads	19	6	2.8033	0.6433	1.5064**
Apm	28	12	9.9474	0.8995	2.4524
Ave	20	5	2.0430	0.5105	0.8736
Bda	32	15	11.0221	0.9093	3.1290***
Bis	33	22	37.7145	0.9735	3.5720
Bon	24	15	21.2308	0.9529	3.5326***
Bra	20	6	3.2204	0.6895	1.3368
Car	24	10	3.0000	0.6667	2.2047***
Cob	30	13	6.9047	0.8552	3.2138***
Coc	24	13	18.4001	0.9457	3.8043*
Cor	24	8	5.8724	0.8297	2.2201
Crc	33	10	3.2592	0.6932	1.2462
Dev	45	16	6.0736	0.8354	1.9879
Eal	25	11	6.5218	0.8467	2.3092***
Eca	24	17	19.7142	0.9493	3.1016
Ega	24	12	6.7318	0.8514	2.9130*
Elb	25	5	3.4483	0.7100	1.6632
Epc	18	13	25.5000	0.9608	3.0131
Esp	25	10	6.9767	0.8567	3.1067***
Esr	24	11	8.1177	0.8768	2.5680**
Est	29	10	5.0123	0.8005	1.6449
Fdf	20	8	5.1351	0.8053	2.4263*
Gar	30	12	7.5000	0.8667	2.3471
Hou	32	10	6.2000	0.8387	1.4133
Lac	33	15	13.2001	0.9242	3.2235*
Lcf	33	9	4.6316	0.7841	1.6023
Lem	33	10	6.0690	0.8352	2.3902**
Ler	20	9	10.0001	0.9000	2.8684
Lig	24	5	3.5845	0.7210	1.2015
Lou	19	10	5.7000	0.8246	2.2456
Mao	19	5	3.2885	0.6959	0.9444
Mau	30	10	6.0416	0.8345	2.4529
Mdb	20	8	3.7255	0.7316	1.3579
Mim	31	15	9.8936	0.8989	3.3081*
Mor	24	11	5.5201	0.8188	1.6240
Mtg	20	8	7.6000	0.8684	2.9211
Mtl	20	7	4.2223	0.7632	1.4689
Ole	20	8	7.0370	0.8579	1.7158
Pan	24	11	6.5714	0.8478	3.2790***
Pdg	32	16	16.5333	0.9395	3.2802**
Pmo	112	18	3.9592	0.7474	1.6855
Sar	24	11	6.4187	0.8442	2.2464**
Seb	33	14	9.4285	0.8939	4.0492***
Sjb	33	10	8.0000	0.8750	3.2235***
Snt	18	7	5.6667	0.8235	2.5948***
Tus	24	7	4.6000	0.7826	1.2790
Var	29	7	5.3421	0.8128	2.0623
Vsg	32	8	5.7012	0.8246	2.8145***
Overall	1339	103	5.9906	0.8260	2.4034
Mean	27.90	10.60	8.2309	0.8254	2.3630
Std	13.67	3.76	6.6426	0.0933	0.8205

11

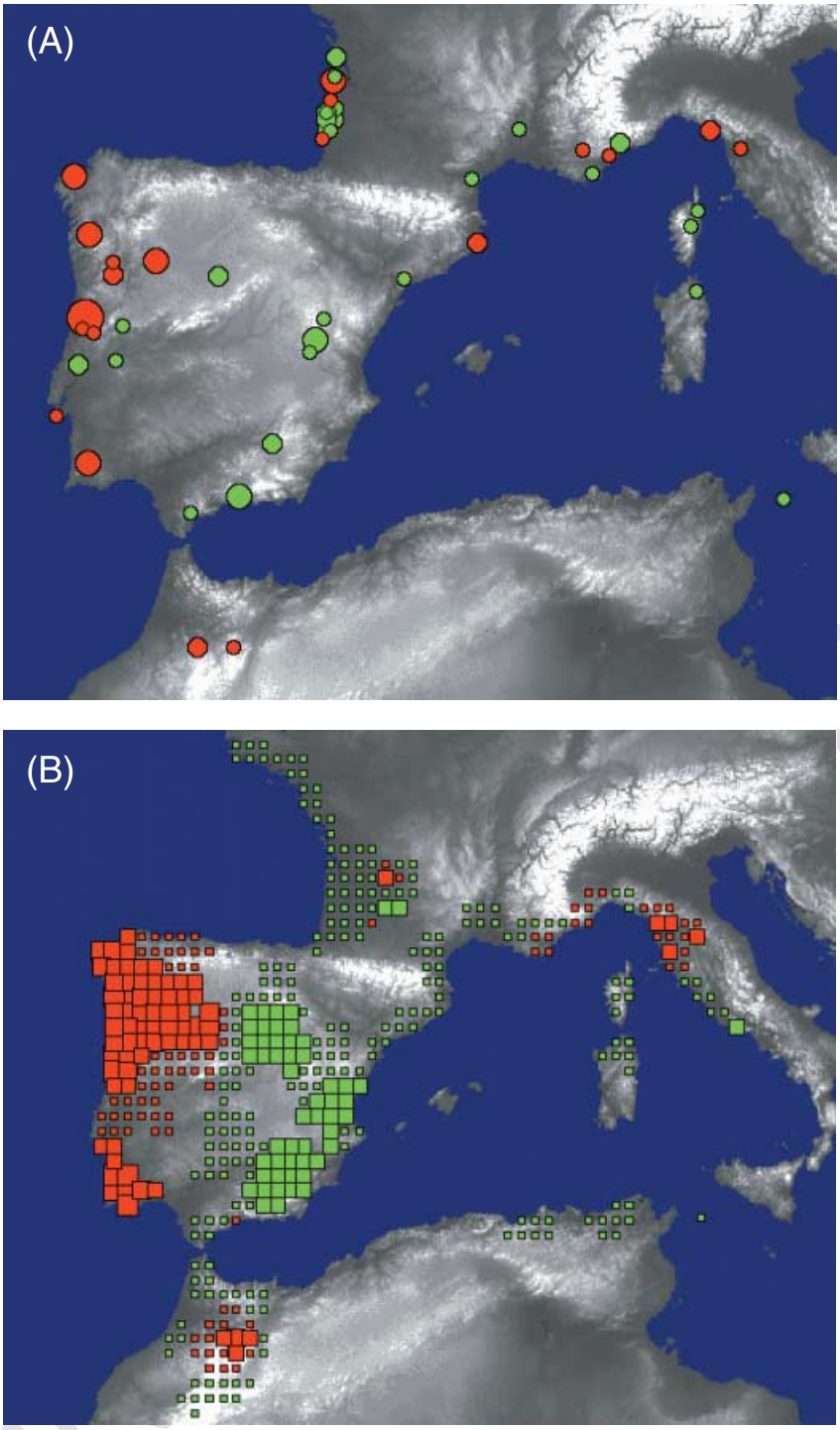

Fig. 2 Geographical distribution of the main diversity parameters within the *Pinus pinaster* range: (A) Within-population unbiased haplotype diversity (H_e, after Nei 1978), standardized to mean = 0 and std = 1, for the 48 stands considered; (B) Haplotypic diversity (H_e — Nei 1987) at the 430 grid points based on 16 spatially interpolated haplotype frequencies (values were standardized to mean = 0 and SD = 1). For both panels, pinpoint dimension is proportional to the parameter value, and point colour represents values above (green) or below (red) the corresponding grand mean.

Colour image

Portugal and the northeastern part of the species range (Fig. S3, supplementary material). Furthermore, some 21 stands (43.75% of the total) showed a distribution of within-population mean pairwise haplotypic differences significantly divergent from Poisson expectations (Table 2), with no recognizable geographical pattern of variation (Fig. S4, supplementary material).

Spatial interpolation of haplotype frequencies and diversity map

To better identify the main genetic gradients across the natural range of *Pinus pinaster*, haplotype frequencies of sampled stands were spatially interpolated over a lattice of 430 grid points covering the range of the species.

Table 3 Goodness-of-fit of haplotypes frequency surfaces based on the 48 sampling-sites raw dataset. A linear regression analysis was carried out, using interpolated values as predictors of the sampled stands haplotypes frequency (i.e. the observed haplotypes) at the same location. For each haplotype, the following regression parameters are reported: the slope and the standard error (SE) of the regression function (no intercept); the *t*-test and its significance; the variance accounted for by the fitting function (adjusted r^2); the mean absolute discrepancy between observed and interpolated values (mean error); and the number of haplotype frequencies falling outside the 5–95% confidence limits of the regression (outliers)

Haplotypes	Slope (*b*)	SE *b*	*t*-value	Pr > |*t*|	Adjusted r^2	Mean Error	Outliers
H01	0.9647	0.0857	11.256	< 0.0001	0.7294	0.0156	1
H02	0.8552	0.0800	10.688	< 0.0001	0.7085	0.0219	2
H03	0.7288	0.1182	6.166	< 0.0001	0.6472	0.0067	2
H04	0.9290	0.1031	9.010	< 0.0001	0.6333	0.003	3
H05	1.3716	0.2375	5.500	< 0.0001	0.4916	0.0126	4
H06	0.6269	0.2363	2.653	0.0109	0.4302	0.0082	2
H07	0.9536	0.1261	7.564	< 0.0001	0.549	0.0011	2
H08	0.8621	0.1336	6.452	< 0.0001	0.4697	0.0039	4
H09	0.6747	0.2066	2.298	0.0261	0.513	0.0021	4
H10	0.8244	0.1819	4.533	< 0.0001	0.4759	0.001	5
H11	0.8040	0.0959	8.387	< 0.0001	0.7649	0.0005	2
H12	0.6959	0.2099	3.315	0.0018	0.5494	0.001	1
H13	0.9804	0.1885	5.200	< 0.0001	0.6233	0.0008	3
H14	1.2192	0.1282	3.715	0.0005	0.5224	0.0013	3
H15	1.3870	0.1344	3.193	0.0025	0.471	0.0016	3
H16	0.9229	0.1239	7.448	< 0.0001	0.7501	0.0006	2

Validation of the haplotype frequency surfaces was carried out using a regression analysis approach. Overall, a good agreement was found between observed and spatially interpolated haplotype frequencies at the same location (Table 3). The mean absolute deviation of the observed haplotype frequencies from their surface-based expected values was 0.0051 ± 0.0064 (range: 0.0005–0.0219). The fairly large variance accounted for by the linear functions fitted (adjusted r^2: 0.4302–0.7649) indicated a low to moderate smoothing of the original sampling data, depending on the haplotype considered and the fine-scale frequency heterogeneity. The number of observed haplotype frequencies falling outside the 5–95% confidence limits of the fitted function varied from 2.08% to 10.41%, with a maximum of 5 for H10 and 4 for H05, H08 and H09.

Haplotypic diversity estimates (H_e) based on spatially interpolated haplotype frequencies (Fig. 2b) showed the lowest H_e values in the westernmost range of the species (Portugal and northern Spain, the 'Atlantic' pool), while the largest diversity was observed for southeastern and central Spain (the 'central' pool).

PCA and evolutionary history of the species

PCA was carried out on the spatially interpolated haplotype frequencies of the 16 most common haplotypes. The first three PCs, which accounted for 35.31%, 19.23% and 11.46% of the variance, respectively (overall 74.29%), were combined into a single synthetic map (Fig. 3). Three main gradients of gene diversity were identified, displayed in Fig. 3 by red, green and blue colours, respectively. Visual inspection of the synthetic map allowed us to hypothesize the existence of three main, independent gene pools across the species' range: (i) a 'southeastern' pool, including southeastern France, Italy, Corsica, Sardinia, Pantelleria, and Morocco; (ii) an 'Atlantic' pool, including all the western area of the Iberian peninsula (Portugal and northwestern Spain); (iii) a 'central' pool, spanning over southeastern Spain. The fourth principal component (Fig. S5, Supplementary material) seems to identify an additional 'pole' of variation centred on SE France and NE Spain (Catalonia), but with fairly low amount of the total variation accounted for (6.89%). Morevover, the existence of two transitional regions may be hypothesized: (iv) an area located in central Spain that seems to be 'transitional' between pools (b) and (c), as detectable in Fig. 3 from its green-blue colour; and (v) the area located in southwestern and central France, showing intermediate PC values for all the three components (represented with pale colours in Fig. 3), and whose genetic composition seems to be a mix of the other groups (as arguable from the colour-patched area of Fig. 3 — see also Fig. S5, Supplementary material).

Cluster analysis and identification of gene zones. In order to define management units and to detect possible gene boundaries within the current species' natural range, cluster analysis was carried out on spatially interpolated haplotype frequencies, obtaining a clear geographical

Fig. 3 Geographical map of the principal component analysis (PCA) results carried out on 16 spatially interpolated haplotype frequencies over a lattice of 430 grid points. Grid points scores (eigenvectors) on the first three principal component (PC) axes (overall variance accounted for: 74.29%) were transformed to a range 0 (no colour) to 255 (full colour), and combined in a single layer using an RGB method (red = PC1, green = PC2, blue = PC3). The relative intensity of the three main colours indicates approximately the position of a given grid point along a main gradient of haplotype frequency variation, as well as their distance from the PC extremes.

Colour image

pattern of eight different gene zones across the natural range of *P. pinaster* (Fig. 4A). The reliability of this classification was further tested by discriminant analysis on grid point frequencies, using the assigned cluster (= gene zone) as grouping variable. All haplotypes showed significant differences in frequency among the eight gene zones detected (data not shown), with an average adjusted $r^2 = 0.6421$ (indicating a fairly good fit of observed values with expectations based on the linear discriminant function obtained). Overall significant differences among zones were supported by multivariate statistics (Roy's greatest root: 12.195; $F_{[16,394]} = 300.31$; $P < 0.0001$). Cross-validation based on the linear discriminant function showed that only 19 out of 430 grid points (4.42%) had a posteriori probability of membership to the assigned gene pool lower than 0.95 (Table 4). In addition, the Ward's minimum-variance dendrogram obtained from the Mahalanobis distance matrix among gene zones (Fig. 4B) suggests that the main subdivision within the species' range occurs in between 'eastern' and 'western' clusters, including (i) western Africa (Morocco), eastern Africa (Pantelleria) and northern Italy (including Corsica and Sardinia); and (ii) all the other gene zones (Iberian peninsula and France), respectively, which agrees fairly well with the main gradients of gene diversity described above. The subdivisions within the 'western' group revealed a gathering of the 'Spanish' gene zones ('Atlantic', 'Central' and 'Eastern Spain', displayed in green colours in Fig. 4A, B), and a 'northern' subgroup ('Northen Spain' and 'France', displayed in blue colours). Further subdivisions within the 'France' zone were not statistically supported (data not shown), despite the large heterogeneity detected for some haplotype frequencies (expected for 'melting pots' — see above).

Genetic divergence among the eight identified gene zones was also tested by AMOVA using mean haplotype frequencies of their grid points: overall genetic divergence among groups (Φ_{CT}) was 0.0117 ($P < 0.001$), while all pairwise genetic differentiation estimates between gene zones were significant though fairly low ($\Phi_{CT} = 0.0129$–0.0871; data not shown). To further assess the reliability of the above subdivisions among gene zones, the 48 sampled stands were reassigned to the eight zones based on their geographical location. AMOVA was applied using raw stand frequencies for all the 103 haplotypes detected, obtaining an overall divergence among the eight groups of populations of $\Phi_{CT} = 0.1036$ ($P < 0.001$).

Finally, seven out of the 48 stands analysed (14.58%) showed significant deviations from their expectations based on the mean spatially interpolated haplotype frequencies of their gene zone (Kolmogorov–Smirnov nonparametric test), and were therefore considered as outliers (Fig. S6, supplementary material). Outlier stands do not seem randomly distributed across the species range: indeed, four out of the seven were

Colour image

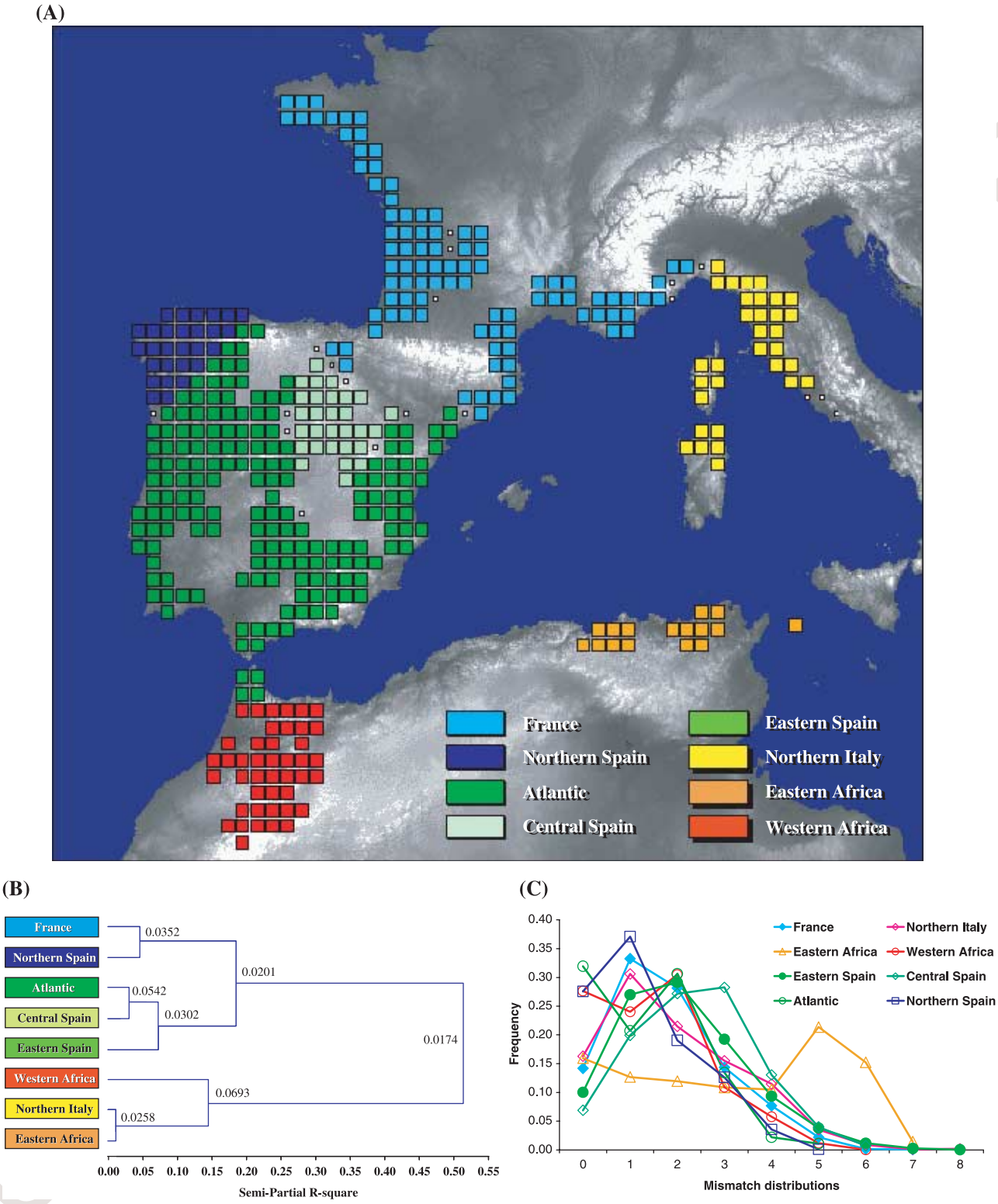

Fig. 4 (A) Geographical delineation of eight gene pools as obtained from cluster analysis based on the 16 spatially interpolated haplotype frequencies. Grid points showing a posteriori probability of membership to the assigned cluster lower than 0.95 (after discriminant analysis) were considered as unclassified (small white squares on the map). (B) Ward's minimum-variance dendrogram based on the Mahalanobis pairwise distance matrix among the eight gene zones identified. Numbers at forks represent the genetic divergence as estimated by Φ_{CT} values. All Φ_{CT} values were highly significant ($P < 0.001$). (C) Chloroplast microsatellite mismatch distribution for the eight gene zones identified.

Table 4 Cross-validation of the 430 grid points (previously classified by cluster analysis) based on the linear discriminant function obtained from the discriminant analysis carried out on the 16 spatially interpolated haplotype frequencies. Grid points were reclassified based on their a posteriori probability of membership to clusters. For each cluster (= gene zone), the number and the proportion (in italics) of grid points correctly ('reassigned') and incorrectly ('unclassified) reclassified is reported. For more details, see text

	Grid Points		
Cluster	Reassigned	Unclassified	Total
Atlantic	102	3	105
	0.9714	*0.0286*	*0.2442*
Northern Spain	27	1	28
	0.9643	*0.0357*	*0.0651*
Central Spain	28	3	31
	0.9032	*0.0968*	*0.0721*
Eastern Spain	71	3	74
	0.9595	*0.0405*	*0.1721*
France	93	6	99
	0.9394	*0.0606*	*0.2302*
Northern Italy	30	2	32
	0.9375	*0.0625*	*0.0744*
Eastern Africa	17	1	18
	0.9444	*0.0556*	*0.0419*
Western Africa	43	0	43
	1.0000	*0.0000*	*0.1000*
Total	411	19	430
	0.9558	*0.0442*	*1.0000*

located in southwestern France (Aquitaine/Landes), probably reflecting admixed populations, while the other three were located in different areas; two of them (northeastern and central Spain) associated with gene zone boundaries, where haplotype frequency surfaces are expected to change sharply in space.

Characterization of gene zones. Allelic frequency and gene diversity estimates for the eight gene zones identified are presented in Table 5. The differences in haplotypic diversity were significant as shown by nonparametric Kruskal–Wallis tests for n_e ($P = 0.005$), H_e ($P = 0.006$) and S_w ($P = 0.022$). Genetic variation, whatever the estimator used, was higher in central and eastern Spain and lower in the gene zones of western Africa, Atlantic and northern Spain. Mismatch distributions were unimodal, reflecting population growth and/or range expansion (Fig. 4C). Population-growth demographic models matched our dataset well ($P > 0.05$ for all gene zones where the model converged — Table 5). Fitted τ values, although difficult to calibrate without external data (for instance, fossil records), showed more ancient expansions in eastern Africa and eastern Spain maritime pine gene zones.

Discussion

The spatial methods used in this investigation allowed extrapolating pinpoint genetic information to neighbouring regions, enabling a better delineation of regions sharing similar genetic make-up of their populations. Moreover, smoothing of sampling data obtained by spatial interpolation allowed the detection of genetic trends at the macro-geographical level by reducing the 'noise' related with sampling errors and/or erratic values. Finally, the spatial methods applied in this work allowed the identification of stands whose genetic make-up is not statistically consistent with local gene zones (outliers). These stands might represent marginal populations or, alternatively, borderlines between gene zones (e.g. central and northeastern Spain), areas of abrupt topography (e.g. Corsica) or the effects of admixture between differentiated native populations or between native populations and material from adjacent stands founded from non-native sources (e.g. Landes, in southwestern France).

Choice of the spatial interpolation method

The method chosen in this work for spatial interpolation of pinpoint genetic data has several advantages in respect to other popular spatial methods (e.g. *kriging* — Matheron 1971; Delfiner 1976) that have been applied in population genetics (Bucci & Vendramin 2000): (i) it is extremely *simple* and *model-free* — weighting the parameter values of the neighbouring sampled points based on their distance from the point to be estimated — and suitable for batch calculations; (ii) it is *adaptive*, i.e. for an unknown point the number of neighbouring sampled points used in the estimation is proportional to the local heterogeneity of frequencies; (iii) as a consequence, it may be *local* (usage of the closest surrounding sampled points only) or *partially local*, depending on the local heterogeneity mentioned above; and (iv) it has proven to provide estimates in good agreement with those obtained with more sophisticated methods (Cavalli-Sforza *et al*. 1994). Spatial analysis by *kriging*, on the other hand, needs to model the function connecting genetic and geographical distance for each parameter considered (i.e. the construction of a variogram), can only be applied to large datasets and is highly sensitive to flaws due to extreme unjustified values in areas where the sampling locations are scarce (Cavalli-Sforza *et al*. 1994). Finally, the methods chosen in our study might apply particularly well to conifer species where high levels of gene flow and ample, continuous populations match the isotropy assumption.

PCA and evolutionary history of the species

The PCA results based on spatially interpolated haplotype frequencies revealed the existence of three main 'poles' of

Table 5 Mean haplotype frequency and average gene diversity (see Table 2) for the eight gene zones identified in this investigation; standard deviations (SD) are in italics. Demographic parameters obtained by fitting the demographic model of Rogers & Harpending (1992) following Schneider & Excoffier (1999) and Excoffier *et al*. (2005) are also given for each gene pool; $\tau = 2\mu t$, where μ is the mutation rate and t is the number of generations since expansion, and P(SSD) is the goodness-of-fit of the demographic model. NC: the demography model did not converge, neither for the gene pool nor for any of the populations within

	Mean haplotype frequency																Genetic diversity			Demography	
Gene Zone	H01	H02	H03	H04	H05	H06	H07	H08	H09	H10	H11	H12	H13	H14	H15	H16	n_e	H_e	S_w	τ	P(SSD)
Atlantic	0.135	0.351	0.040	0.113	0.006	0.003	0.047	0.089	0.028	0.020	0.003	0.007	0.029	0.002	0.002	0.014	5.1958	0.7653	0.3711	2.095	0.3020
	0.041	*0.087*	*0.016*	*0.036*	*0.005*	*0.003*	*0.015*	*0.042*	*0.019*	*0.011*	*0.003*	*0.005*	*0.012*	*0.002*	*0.002*	*0.009*	*2.3539*	*0.1151*	*0.2677*		
Central Spain	0.183	0.141	0.057	0.077	0.012	0.012	0.045	0.022	0.012	0.049	0.010	0.027	0.050	0.005	0.003	0.010	18.4001	0.9457	1.0616	NC	NC
	0.013	*0.036*	*0.009*	*0.006*	*0.005*	*0.007*	*0.003*	*0.008*	*0.004*	*0.013*	*0.005*	*0.006*	*0.013*	*0.003*	*0.001*	*0.005*	—	—	—		
Eastern Spain	0.204	0.136	0.037	0.079	0.019	0.017	0.062	0.022	0.013	0.016	0.011	0.012	0.039	0.006	0.003	0.048	13.5419	0.8992	0.7702	2.445†	—
	0.033	*0.044*	*0.020*	*0.020*	*0.009*	*0.016*	*0.020*	*0.014*	*0.006*	*0.009*	*0.006*	*0.006*	*0.013*	*0.003*	*0.001*	*0.035*	*8.2502*	*0.0525*	*0.1851*		
Northern Spain	0.299	0.295	0.027	0.060	0.016	0.004	0.034	0.063	0.010	0.019	0.004	0.005	0.015	0.002	0.022	0.021	3.1443	0.6813	0.3319	1.280	0.1610
	0.097	*0.112*	*0.015*	*0.021*	*0.010*	*0.003*	*0.010*	*0.024*	*0.006*	*0.010*	*0.003*	*0.003*	*0.007*	*0.002*	*0.021*	*0.011*	*0.2040*	*0.0206*	*0.3387*		
France	0.231	0.145	0.121	0.042	0.036	0.032	0.022	0.012	0.012	0.032	0.034	0.032	0.011	0.021	0.008	0.009	9.2552	0.8510	0.6365	2.147 †	—
	0.028	*0.029*	*0.030*	*0.013*	*0.015*	*0.021*	*0.007*	*0.007*	*0.003*	*0.008*	*0.012*	*0.006*	*0.006*	*0.008*	*0.003*	*0.004*	*7.6721*	*0.0721*	*0.3434*		
Northern Italy	0.206	0.222	0.046	0.038	0.049	0.151	0.023	0.003	0.070	0.006	0.024	0.008	0.003	0.010	0.009	0.004	5.3098	0.8026	0.3547	1.984	0.1750
	0.017	*0.057*	*0.032*	*0.015*	*0.018*	*0.041*	*0.007*	*0.003*	*0.032*	*0.006*	*0.006*	*0.007*	*0.003*	*0.006*	*0.007*	*0.003*	*1.1843*	*0.0515*	*0.2153*		
Eastern Africa	0.201	0.152	0.072	0.053	0.062	0.056	0.036	0.018	0.024	0.020	0.016	0.014	0.018	0.019	0.014	0.014	6.5714	0.8478	0.5018	5.321	0.4580
	0.004	*0.001*	*0.002*	*0.003*	*0.010*	*0.006*	*0.004*	*0.002*	*0.003*	*0.001*	*0.000*	*0.000*	*0.002*	*0.001*	*0.003*	*0.002*	—	—	—		
Western Africa	0.145	0.290	0.018	0.060	0.077	0.005	0.066	0.061	0.024	0.010	0.007	0.008	0.014	0.023	0.009	0.023	4.7397	0.7831	0.2742	1.939	0.3340
	0.011	*0.060*	*0.006*	*0.011*	*0.043*	*0.003*	*0.014*	*0.013*	*0.008*	*0.002*	*0.001*	*0.001*	*0.004*	*0.013*	*0.004*	*0.010*	*1.1037*	*0.0505*	*0.1223*		
Overall	0.192	0.225	0.057	0.071	0.029	0.026	0.042	0.041	0.022	0.022	0.015	0.015	0.023	0.011	0.007	0.019	8.2309	0.8254	0.5441		
	0.061	0.110	0.042	0.035	0.028	0.042	0.021	0.040	0.021	0.014	0.014	0.012	0.016	0.010	0.008	0.021	6.6426	0.0933	0.3227		

†For France and east of Spain gene pools the algorithm did not converge and average values from single populations within the gene pool are provided instead; n_e, effective number of haplotypes; H_e, haplotypic diversity; S_w, within-stand haplotypic variance (Slatkin 1995).

genetic variation within the range of *Pinus pinaster* , which may be hypothesized to correspond to three distinct gene pools, i.e. populations or regions likely to share the same evolutionary history in terms of origin from common glacial refugia and colonization along migration routes. Geographical areas showing intermediate values of the first three components may be interpreted here as 'transitional' regions among the main pools (e.g. central Spain), where admixture of divergent lineages from separate refugia might occur (Petit *et al*. 2003).

However, based on our results, some exception to the above scenario may be figured out. The accumulation of 'outliers' in the Aquitaine region (Landes, southwestern France), previously reported as a 'hotspot' of haplotypic variation for the species (Vendramin *et al*. 1998), seems to support the existence of a mix of stands from other pools ('melting pot'). Moreover, lower-rank PCs suggest the existence of an additional 'pole' of haplotypic variation (with fairly low amounts of the total variation accounted for) spanning from the French Mediterranean shoreline up to the northeast of Spain; although the region mentioned still retains intermediate characteristics (as for higher-rank PCs) in between 'western' and 'eastern' gene pools. A refined discussion on the admixture of different gene pools or about hybrid zones in maritime pine is out of the scope of this paper, in that both organelle markers are not well suited for this kind of study and current sampling in contact areas is incomplete and/or represented by only a few populations.

The above range-wide picture overlaps with that previously reported by other authors for the same species, with some exceptions. The existence of a main genetic subdivision between eastern and western gene pools have been reported previously by Burban & Petit (2003) based on mitotypes. Moreover, these authors hypothesized the existence of a third independent gene pool in Morocco, while evidence reported here supports its relationship to a wider 'southeastern' gene pool, including populations from northern Africa, Pantelleria, Sardinia, Corsica and northern Italy. In addition, the Mediterranean coast of France (and northeastern Spain) was classified by the above authors as belonging to the eastern gene pool, while in our work the whole region is split by the Rhône valley into 'eastern' (Provence and Cote d'Azur, characterized by reddish colour in the PC synthetic map) and 'western' (Narbonnaise and Catalonia, white-greenish colours) gene pools. Additional sampling sites for this region are needed to shed light on the above puzzling discordance.

Baradat & Marpeau (1988) hypothesized the existence of glacial refugia in southwestern Iberia peninsula, from which both the northwards recolonization of Spain, France and Italy, and the southwards recolonization of northern Africa, might have taken place (see also Bahrman *et al*. 1994). Based on evidences reported here, the above scenario may be definitely rejected, while the occurrence of an 'African' recolonization pathway eastwards to Italy (and maybe westwards to Morocco, Vendramin *et al*. 1998) seems more plausible. In addition, our results support the existence of glacial refugia in southeastern Spain and the Atlantic coast of Portugal, as suggested by Salvador *et al*. (2000) and Ribeiro *et al*. (2001), respectively, from which northward recolonization of central Spain and the northwestern range of the species might have taken place.

Map of haplotypic diversity

The application of spatial interpolation methods produced a range-wide map of the haplotypic diversity for *P. pinaster*, enabling the identification of 'reservoirs' of genetic variation for the species, along with regions showing a more impoverished genetic make-up. It has to be remarked here that spatial interpolation is expected to smoothen out small-scale stochastic effects on local gene frequencies (Bucci & Vendramin 2000), and consequently the derived diversity map is expected to better describe the 'average' genetic diversity across the species range, in that it represents the probability of detecting a particular level of diversity by sampling stands at random in a given unsampled region. However, the existence of particular areas or sites showing large diversity at local scale should not be excluded.

In this study, maximum haplotypic diversity was found in southeastern and central Spain, which therefore may be considered as 'hotspot' of intraspecific biodiversity. As for the latter region (i.e. the Castilian Plateau in central Spain), it may be hypothesized that large diversity may stem from admixture between the 'Atlantic' and the 'central' pools described above, and from the multiple and variable (unsampled) maritime pine populations inhabiting surrounding mountains. On the other hand, low to very low haplotypic diversity was detected for: (i) a wide area located in the western Iberian peninsula (Portugal and northwestern Spain), which also showed the lowest values for number of private haplotypes and within-population haplotypic variance; and (ii) Morocco, whose populations are typically small and fragmented and also severely threatened by overexploitation and lack of natural regeneration (Wahid *et al*. 2004 and references therein).

Delineation and characterization of gene zones

For forest tree species, breeding zones delineation within a continuous range of distribution is usually traced out based on climatic, edaphic, genetic, adaptive and management criteria (Westfall & Conkle 1992; Yang *et al*. 1998; Bucci & Vendramin 2000; Hamann *et al*. 2000; González-Martínez *et al*. 2004). In this investigation, we have applied a method based on the spatial interpolation of chloroplast microsatellite

haplotype frequencies to delineate genetically homogeneous zones to provide a genetic 'baseline' for *P. pinaster*, to be combined with climatic, adaptive traits-related and management information in the definition of conservation and breeding zones for the species studied.

Using cluster and discriminant analysis, we were able to identify genetic subdivisions across the species' range, both among and within the gene pools described above. Indeed, the whole species' range has been split into eight different gene zones, whose genetic relationships seem to be related, to some extent, with the evolutionary history of the species. Moreover, by fitting demographic models, more ancient expansions in eastern Africa and southeastern Spain maritime pine gene zones were demonstrated, which supports the phylogeographical inferences reported above.

The delineation of gene zones may therefore be considered as the first step to design management units (MUs) *sensu* Moritz (1994) within the natural range of *P. pinaster* (see also González-Martínez *et al*. 2004). However, it is worth mentioning some general considerations on possible flaws related to the approach adopted here: First, the robustness of the gene zone delineation is a function of the number and regular distribution in space of sampling sites used for frequency-surfaces construction (Cavalli-Sforza *et al*. 1994). Scarcely sampled regions may provide more 'shaky' estimates, being more prone to sampling error and/or erratic values within the raw dataset. In our work, areas located in northern Africa (Rif and High Atlas in Morocco and Tunisia/Algeria), central Spain or inner and northwestern France have a low density of sampling sites, and therefore their genetic parameters should be considered with caution. Second, the algorithm used in this investigation for spatial interpolation of sampling data has a smoothing effect on haplotype frequencies, making changes in haplotype surfaces less steep (though still statistically significant) than expected based on raw stand data. As a consequence, divergence values among grid point clusters are expected to be lower than sampling stand clusters. Based on the above consideration, it is strongly recommended that sampling activities aimed at conservation/breeding purposes be located near the central area of each gene zone identified in this investigation. Third, local shift of gene frequencies due to processes not related to evolutionary history (e.g. local selection, introgression from allochthonous gene pools of the species etc.) may result in statistically significant discontinuities within the parent 'gene pool', leading to the identification of independent gene zones after cluster and discriminant analysis (as it may be hypothesized for northwestern Spain gene zone in this work). On the other hand, smooth variation of frequencies in space between adjacent gene zones and/or large-scale admixtures of different lineages may lead to failures in the detection of extant genetic discontinuities (as it seems the case for the southeastern France region, clustering with the 'France' gene zone, though an additional, low-rank pole of genetic variation was identified in the same area).

For the above reasons, the method used needs to be confirmed by further sampling (mainly in underrepresented or unresolved regions) before its application as an operational tool for supporting the decision-making process in the delineation of conservation areas or species management units. Here, our main goals were to provide a preliminary baseline for gene-zone delineation and to identify putative 'transition' areas in order to develop detailed studies in the future.

Acknowledgements

The technical work of Roberta Pastorelli is highly appreciated. This work was supported by the European Union (INCO project FORADAPT 'Global, physiological and molecular responses to climatic stresses of three Mediterranean conifers', contract ERBIC 18CT 970 200). The work of Santiago C. González-Martínez was supported by the 'Ramón y Cajal' fellowship RC02-2941 and the projects CC95-0097 (INIA/Ministerio de Medio Ambiente) and REPROFOR (AGL2005-07440-C02-01, Ministerio de Educación y Ciencia).

Supplementary material

The supplementary material is available from
http://www.blackwellpublishing.com/products/journals/suppmat/MEC/MEC3275/MEC3275sm.htm

Fig. S1 Examples of the geographical distribution of frequencies for six of the 16 haplotypes analysed in this work. Point size is proportional to sampling frequencies, rescaled to a range of 0–1.

Fig. S2 Geographical distribution of the number of private haplotypes (standardized to mean = 0 and SD = 1) for the 48 stands analysed. Pinpoint dimension is proportional to the absolute parameter value, while point colour represents values above (white) or below (grey) the grand mean.

Fig. S3 Geographical distribution of the within-population haplotypic variance (S_w, after Slatkin 1995, standardized to mean = 0 and SD = 1) for the 48 stands analysed. Pinpoint dimension is proportional to the absolute parameter value, while point colour represents values above (white) or below (grey) the grand mean.

Fig. S4 Geographical distribution of the stands showing significant deviation of the within-population mean number of haplotypic differences (x_{ij}) from the expected Poisson distribution. Dark-grey large points, $P < 0.001$; medium-grey medium-sized points, $P < 0.01$; light-grey small points, $P < 0.05$; white smallest points, n.s.

Fig. S5 Results of the principal component analysis (PCA) carried out on spatially interpolated haplotype frequencies of the 430 grid points covering the current range of *Pinus pinaster*. Figures in parentheses refer to the proportion of the total haplotype-frequency variance accounted for by each component.

Fig. S6 Geographical distribution of sampled stands showing significant deviation from gene zones' haplotypes means (after Kolmogorov–Smirnov nonparametric test). White points, n.s.; light grey points, $P < 0.05$; dark grey points, $P < 0.01$.

Table S1 List of the 48 populations sampled in this investigation. Geographical coordinates are in decimal degrees (negative values: degrees of longitude west of Greenwich)

References

Acheré V, Favre Jm, Besnard G, Jeandroz S (2005) Genomic organization of molecular differentiation in Norway spruce (*Picea abies*). *Molecular Ecology*, **14**, 3191–3201.

Alía R, Moro J, Denis JB (1997) Performance of *Pinus pinaster* provenances in Spain: interpretation of the genotype by environment interaction. *Canadian Journal of Forest Research*, **27**, 1548–1559.

Bahrman N, Zivy M, Damerval C, Baradat P (1994) Organisation of the variability of abundant proteins in seven geographical origins of maritime pine (*Pinus pinaster* Ait.). *Theoretical and Applied Genetics*, **88**, 407–411.

15

Baradat P, Marpeau-Bezard A (1988) *Le pin maritime Pinus pinaster Ait.: biologie et génétique des terpénes pour la connaissance et l'amélioration de l'espéce*. University of Bordeaux-I, Bordeaux.

Barbujani G (1988) Diversity of some gene frequencies in European and Asian populations. IV. Genetic population structure assessed by the variograms. *Annals of Human Genetics*, **52**, 215–225.

Bozhko M, Riegel R, Schubert R, Muller-Starck G (2003) A cyclophilin gene marker confirming geographical differentiation of Norway spruce populations and indicating viability response on excess soil-born salinity. *Molecular Ecology*, **12**, 3147–3155.

Brendel O, Pot D, Plomion C, Rozenberg P, Guehl JM (2002) Genetic parameters and QTL analysis of $\delta^{13}C$ and ring width in maritime pine. *Plant, Cell and Environment*, **25**, 945–953.

Bucci G, Vendramin GG (2000) Delineation of genetic zones in the European Norway spruce natural range: preliminary evidence. *Molecular Ecology*, **9**, 923–934.

Burban C, Petit RJ (2003) Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. *Molecular Ecology*, **12**, 1487–1495.

Burban C, Petit RJ, Carcreff E, Jactel H (1999) Rangewide variation of the maritime pine bast scale *Matsucoccus feytaudi* Duc. (*Homoptera*: *Matsucoccidae*) in relation to the genetic structure of its host. *Molecular Ecology*, **8**, 1593–1602.

Carrión JS, Van Geel B (1999) Fine-resolution Upper Weichselian and Holocene palynological record from Navarrés (Valencia, Spain) and a discussion about factors of Mediterranean forest succession. *Review Palaeobotany Palynology*, **106**, 209–236.

Carrión JS, Yll EI, Walker MJ, Legaz AJ, Chain C, Lopez A (2003) Glacial refugia of temperate, Mediterranean and Ibero-North African flora in south-eastern Spain: new evidence from cave pollen at two Neanderthal man sites. *Global Ecology and Biogeography*, **12**, 119–129.

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) *History and Geography of Human Genes*, p. 1024. Princeton University Press, Princeton, NY.

Chambel R (2006) *Variabilidad Adaptativa y Plasticidad Fenotípica en Procedencias de Pinos Ibéricos*. PhD Dissertation, UPM, Madrid.

Covassi M, Binazzi A (1992) Primi focolai di *Matsococcus feytaudi* Duc. nella Liguria orientale. *Redia*, **75**, 453–466.

Davis JC (1973) *Statistics and Data Analysis in Geology*. John Wiley & Sons, New York, NY.

16

Derory J, Mariette S, Gonzaléz-Martínez SC *et al.* (2002) What can nuclear microsatellites tell us about maritime pine genetic resources conservation and provenances certification strategies? *Annals of Forest Science*, **59**, 699–708.

Desprez LM, Baradat P (1991) Variation in susceptibility to twisting rust of maritime pine. *Annales Des Science Forestières*, **48**, 497–511.

Devy-Vareta N (1988) La question du reboisement au Portugal, un processus de longue durée. *Revue Géographique Des Pyrénées et Du Sud-Ouest*, **59**, 159–186.

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus*, **12**, 13–15.

Excoffier L (2004) Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Molecular Ecology*, **13**, 853–864.

Excoffier L, Laval G, Schneider S (2005) Arlequin version 3.0: An integrated software package for population genetics data analysis, p. 111. [online] URL: http://cmpg.unibe.ch/software/arlequin3.

Gómez A, Vendramin GG, González-Martínez SC, 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. *Diversity and Distributions*, **11**, 257–263.

Gömöry D, Hynek V, Paule L (1998) Delineation of seed zones for European beech (*Fagus sylvatica* L.) in the Czech Republic based on isozyme markers. *Annales des Science Forestières*, **55**, 425–436.

González-Martínez SC, Agúndez D, Alía R, Salvador L, Gil L (2001) Geographical variation of gene diversity of *Pinus pinaster* Ait. in the Iberian Peninsula. In: *Genetic Response of Forest Systems to Changing Environmental Conditions* (eds Schubert R, Müller-Starck G), pp. 161–171. Kluwer Academic Press, Dordrecht.

González-Martínez SC, Mariette S, Ribeiro MM *et al.* (2004) Genetic resources in maritime pine (*Pinus pinaster* Ait.): patterns of differentiation and correlation between molecular and quantitative measures of genetic variation. *Forest Ecology and Management*, **197**, 103–115.

González-Martínez SC, Gil L, Alía R (2005) Genetic diversity estimates of *Pinus pinaster* in the Iberian Peninsula: a comparison of allozymes and quantitative traits. *Investigación Agraria: Sistemas y Recursos Forestales*, **14**, 3–12.

Green DM (2005) Designatable units for status assessment of endangered species. *Conservation Biology*, **19**, 1813–1820.

Guenni L, Hutchinson MF (1998) Spatial interpolation of the parameters of a rainfall model from ground-based data. *Journal of Hydrology*, **213**, 355–347.

Hamann A, Koshy MP, Namkoong G, Ying CC (2000) Genotype × environment interactions in *Alnus rubra*: developing seed zones and seed-transfer guidelines with spatial statistics and GIS. *Forest Ecology and Management*, **136**, 107–119.

Higdon D (1998) A process-convolution approach to modeling temperatures in the north Atlantic ocean. *Environmental and Ecological Statistics*, **5**, 173–190.

Lagerkrantz U, Ryman N (1990) Genetic structure of Norway Spruce (*Picea abies*): concordance of morphological and allozymic variation. *Evolution*, **44**, 38–53.

Le Corre V, Roussel G, Zanetto A, Kremer A (1998) Geographical structure of gene diversity in *Quercus petraea* (Matt.) Liebl. III. Patterns of variation identified by geostatistical analyses. *Heredity*, **80**, 464–473.

Loustau D, Crepeau S, Guye MG, Sartore M, Saur E (1995) Growth and water relations of three geographically separate origins of maritime pine (*Pinus pinaster*) under saline conditions. *Tree Physiology*, **15**, 569–576.

Matheron G (1971) *The Theory of Regionalized Variables and its Applications*. Ecole Nationale Supérieure des Mines, Paris.

Monastiez P, Goulard M, Charmet G (1999) Geostatistics for spatial genetic structures: study of wild populations of perennial ryegrass. *Theoretical and Applied Genetics*, **88**, 33–41.

Moritz C (1994) Defining 'Evolutionary Significant Units' for conservation. *Trends in Ecology and Evolution*, **9**, 373–375.

Nalder IA, Wein RW (1998) Spatial interpolation of climatic Normals: test of a new method in the Canadian boreal forest. *Agricultural and Forest Meteorology*, **92**, 211–225.

Nanos N, González-Martínez SC, Bravo F (2004) Studying within-stand structure and dynamics with geostatistical and molecular marker tools. *Forest Ecology and Management*, **189**, 223–240.

Nanos N, Pardo F, Nager JA, Pardos JA, Gil L (2005) Using multivariate factorial kriging for multiscale ordination: a case study. *Canadian Journal of Forest Research*, **35**, 2860–2874.

Navascués M, Vaxevanidou Z, González-Martínez SC, Climent J, Gil L (2006) Chloroplast microsatellites reveal colonisation and metapopulation dynamics in the Canary Island pine. *Molecular Ecology*, **15**, 2691–2698.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York, NY.

Newton AC, Allnutt TR, Gillies ACM, Lowe AJ, Ennos RA (1999) Molecular phylogeography, intrapecific variation and the conservation of tree species. *Trends in Ecology and Evolution*, **14**, 140–145.

Nguyen A, Lamant D (1989) Variation in growth and osmotic regulation of roots of water-stressed maritime pine (*Pinus pinaster* Ait.) provenances. *Tree Physiology*, **5**, 123–133.

Nielsen R, Tarpy DR, Reeve HK (2003) Estimating effective paternity number in social insects and the effective number of alleles in a population. *Molecular Ecology*, **12**, 3157–3164.

Pamilo P (1993) Polyandry and allele frequency differences between the sexes in the ant *Formica aquilonia*. *Heredity*, **70**, 472–480.

Petit RJ, Pineau E, Demesure B *et al.* (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences USA*, **94**, 9996–10001.

Petit RJ, Aguinagalde I, de Beaulieu JL *et al.* (2003) Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.

Piazza A, Menozzi P, Cavalli-Sforza LL (1981) The making and testing of gene-frequency maps. *Biometrics*, **37**, 635–659.

Pot D, McMillan L, Echt C *et al.* (2005) Nucleotide variation in genes involved in wood formation in two pine species. *New Phytologist*, **167**, 101–112.

Ray N, Currat M, Excoffier L (2003) Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution*, **20**, 76–86.

Ribeiro MM, Plomion C, Petit R, Vendramin GG, Szmidt AE (2001) Variation of chloroplast simple-sequence repeats in Portuguese maritime pine (*Pinus pinaster* Ait.). *Theoretical and Applied Genetics*, **102**, 97–103.

Ribeiro MM, Garnier-Géré PH, González-Martínez SC *et al.* (2006) Chapter 4.3 Genetic Diversity. In: *Maritime Pine Monograph* (ed. Timbal J). INRA. (in press).

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

Salvador L, Alía R, Agúndez D, Gil L (2000) Genetic variation and migration pathways of maritime pine (*Pinus pinaster* Ait) in the Iberian peninsula. *Theoretical and Applied Genetics*, **100**, 89–95.

Schneider S, Excoffier L (1999) Estimation of demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: Application to human mitochondrial DNA. *Genetics*, **152**, 1079–1089.

Shepard D (1968) A two-dimensional interpolation function for irregularly spaced data. *Proceedings of the 1968 ACM National Conference*, pp. 517–524. Brandon Systems, Princeton, New Jersey.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129**, 555–562.

Taylor BL, Dizon AE (1999) First policy then science: why a management unit based solely on genetic criteria cannot work. *Molecular Ecology*, **8**, S11–S16.

Teixeira C (1945) Subsídios para a história evolutiva do pinheiro dentro da flora portuguesa. *Boletim da Sociedade Broteriana*, **19**, 209–221.

Vendramin GG, Lelli L, Rossi P, Morgante M (1996) A set of primers for the amplification of 20 chloroplast microsatellite in *Pinaceae*. *Molecular Ecology*, **5**, 595–598.

Vendramin GG, Anzidei M, Madaghiele A, Bucci G (1998) Distribution of genetic diversity in *Pinus pinaster* Ait. as revealed by chloroplast microsatellites. *Theoretical and Applied Genetics*, **97**, 456–463.

Vidakovic M (1991) *Conifers: Morphology and Variation*. Graficki Zavod Hravtske, Zagreb.

Wahid N, González-Martínez SC, El Hadrami I, Boulli A (2004) Genetic structure and variability of the Moroccan natural populations of maritime pine (*Pinus pinaster* Aiton). *Silvae Genetica*, **53**, 93–99.

Westfall RD, Conkle MT (1992) Allozyme markers in breeding zones designation. In: *Population Genetics of Forest Trees* (eds Adams WT, Strauss SH, Copes DL, Griffin AR), pp. 279–309. Kluwer Academic Publisher, Dordrecht.

Yang J-C, Cheng C-C, Kung FH (1998) Defining *Cryptomeria* seed sources useful for Taiwan by superimposing probabilities of good provenance results over climatic data maps. *Silvae Genetica*, **47**, 190–196.

Ziegenhagen B, Kormutak A, Shauerte M, Scholz F (1995) Restriction site polymophism in chloroplast DNA of silver fir (*Abies alba* Mill.). *Forest Genetics*, **2**, 99–107.

Author Query Form

Journal: Molecular Ecology

Article: mec_3275.fm

Dear Author,

During the copy-editing of your paper, the following queries arose. Please respond to these by marking up your proofs with the necessary changes/additions. Please write your answers on the query sheet if there is insufficient space on the page proofs. Please write clearly and follow the conventions shown on the attached corrections sheet. If returning the proof by fax do not write too close to the paper's edge. Please remember that illegible mark-ups may delay publication.

Many thanks for your assistance.

No.	Query	Remarks
1	Lousteau et al. 1995 has been changed to Loustau et al. 1995 so that this citation matches the list.	
2	Taylor 1999 has been changed to Taylor & Dizon 1999 so that this citation matches the list.	
3	Did Acheré *et al.* 2005 reveal hints **of** differential local adaptaion, or do their findings hint **at** differential local adaptaion? Please check and edit.	
4	Westfall & Conckle 1992 has been changed to Westfall & Conkle 1992 so that this citation matches the list.	
5	Gomory et al. 1998 has been changed to Gömöry et al. 1998 so that this citation matches the list.	
6	Ziegenhagen *et al.* 1993 has been changed to Ziegenhagen *et al.* 1995 so that this citation matches the list.	
7	Please check the sentence starting with 'Five chloroplast...'. I added 'of' before 'high polymorphism. Is that correct? the sense still isn't very clear, I'm afraid.	
8	Is 'sdt' Standard Deviation?	
9	Petit et al. 2004 has been changed to Petit et al. 2003 so that this citation matches the list.	
10	Hudson & Slatkin 1991 has not been found in the list.	
11	Nei 1978 has not been found in the list.	
12	Delfiner 1976 has not been found in the list.	

No.	Query	Remarks
13	'Pole' or 'pool' in 'low-rank pole of genetic variation'? Please check and edit.	
14	Slatkin 1994 has been changed to Slatkin 1995 so that this citation matches the list .	
15	A letter 'a' has been deleted from the year 1988. Please confirm.	
16	Davis 1973 has not been found in the text	
17	Please insert the city of publication (poblisher INRA) for reference Ribeiro 2006. Is this Paris?	
18	Please supply a paragraph about the research interests of the authors.	

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

Instruction to printer	*Textual mark*	*Marginal mark*
Leave unchanged	··· under matter to remain	
Insert in text the matter indicated in the margin		New matter followed by or
Delete	through single character, rule or underline or through all characters to be deleted	or
Substitute character or substitute part of one or more word(s)	through letter or through characters	new character or new characters
Change to italics	under matter to be changed	
Change to capitals	under matter to be changed	
Change to small capitals	under matter to be changed	
Change to bold type	under matter to be changed	
Change to bold italic	under matter to be changed	
Change to lower case	Encircle matter to be changed	
Change italic to upright type	(As above)	
Change bold to non-bold type	(As above)	
Insert 'superior' character	through character or where required	or under character e.g. or
Insert 'inferior' character	(As above)	over character e.g.
Insert full stop	(As above)	
Insert comma	(As above)	
Insert single quotation marks	(As above)	or and/or or
Insert double quotation marks	(As above)	or and/or or
Insert hyphen	(As above)	
Start new paragraph		
No new paragraph		
Transpose		
Close up	linking characters	
Insert or substitute space between characters or words	through character or where required	
Reduce space between characters or words	between characters or words affected	