Genetic variation and divergence in Scots pine (Pinus sylvestris L.) within its natural range in Italy

Genetic variation and divergence in Scors pine *(Pinus sylvestris L.* within its natural range in Italy

by P. Belletti, D. Ferrazzini, A. Piotti, I. Monteleone, F. Ducci

[European Journal of Forest Research, 131 (4): 1127-1138, 2012, DOI 10.1007/s10342-011-0584-3]

Genetic variation and divergence in Scots pine (*Pinus sylvestris* L.) within its natural range in Italy

P. Belletti[1,*], D. Ferrazzini[1], A. Piotti[2], I. Monteleone[1], F. Ducci[3]

[1] University of Turin, DIVAPRA Plant Genetics and Breeding, Grugliasco, Italy
[2] University of Parma, Department of Environmental Sciences, Parma, Italy
[3] Agricultural Research Council, Institute of Forestry, Arezzo, Italy

** Corresponding Author (via Leonardo da Vinci 44, 10095 Grugliasco, Italy - phone +39 011 6708804, fax +39 011 2368804, E-mail address: piero.belletti@unito.it)*

Abstract

Twenty-one populations of Scots pine sampled over the entire Italian range of the species were analysed for genetic variation scored at 9 nuclear SSR markers. The main aim of the work was to find genetic features useful for conservation management, namely allelic composition, gene diversity and differentiation.

High levels of intra-population variability were scored. The only population sampled in the Apennines gave the lowest values, confirming the genetic erosion undergone in the Scots pine remnants in this area. A low level of genetic variability was also scored for populations from the Po valley and hills of Piedmont. Most genetic diversity was found within-populations, while only a small amount occurred among them (F_{ST} = 0.058). Both Bayesian clustering and sPCA analysis showed a East-West subdivision, notwithstanding the unclear position of populations from the Po valley. The population from the Apennines was always clearly separated from the others.

The results are discussed in terms of post-glacial recolonisation, as well as for defining the regions of provenance of Scots pine in Italy. The management of genetic resources could benefit from the identification of genetically homogeneous regions, thereby avoiding the use of non-local reproductive material for plantations, which is well known as one of the most important reason for failure of reforestation.

Key words

Genetic differentiation, genetic variation, glacial refugia, regions of provenance, Scots pine, SSR markers

Introduction

Genetic erosion is one of the most serious threats for the survival of forest ecosystems worldwide. A high level of variability is essential to supply populations with strong adaptability. This is particularly important for forest species, which consist of individuals with long life-cycles and no possibility of migrating to more favourable sites (Palmberg-Lerche 2001, Toro and Caballero 2005). Genetic erosion can be strongly enhanced by habitat fragmentation and marginality in the species' range (Jump and Peñuelas 2006, Eckert et al. 2008). However, the expected 'genetic signal' depending on increasing isolation and decreasing population size is often undetected, as fragmentation may have occurred relatively recently and fragments may contain large remnant populations and the longevity of many tree species can delay the loss of genetic variability (Kramer et al. 2008). In addition, the historically underestimated long distance dispersal capability of forest trees can also maintain a high connectivity among widely isolated stands (Robledo-Arnuncio and Gil 2005, Nathan 2006, Williams 2010). Conversely, climate warming can strengthen the loss of genetic variability due to fragmentation, causing a marked decline in growth and survival of marginal southern populations of temperate tree species, as demonstrated in European *Pinus sylvestris* populations by Rubiales et al. (2008) and Reich and Oleksyn (2008).

Knowledge of the level and distribution of genetic variation is of the utmost importance in providing information for the conservation and the management of genetic resources. Furthermore, genetic analysis based on molecular markers can increase our understanding of the historical processes that led to the present distribution of a species (Petit et al. 2003). These markers can provide us with appropriate means to obtain information on the genetic structure of populations, as well as to analyse the distribution of within-species variability (Pautasso 2009). Such data provide important insights for preservation and restoration programs, indicating areas of high genetic diversity and geographic limits for seed collection, and delimiting the scale at which conservation should be planned (Escudero et al. 2003).

Scots pine (*Pinus sylvestris* L.) is the most widespread European conifer tree, and its natural range extends from the arctic circle in Scandinavia down to southern Spain and central Italy, and from western Scotland to eastern Siberia. In southern Europe and Asia Minor, isolated occurrences are confined to the

mountain zone (up to 2200 m in altitude in the Balkans and Spain, and 2700 m in the Caucasus). In Italy the species is widely spread throughout the Alps, and some relic populations can be found in the northern Apennines, in the hilly areas of Piedmont (north-western Italy) and in the upper part of the Po valley (Pignatti 1982). The present distribution is highly influenced by human activities. In the Alps Scots pine has often been substituted by other species, namely Norway spruce and black pine, while the other populations are nowadays regressing, mainly due to recruitment limitation, competition with other forest trees and as a consequence of rural depopulation and the abandonment of wood management (Camerano et al. 2008).

The complex biogeographic history of Scots pine in Europe has been extensively studied (Sinclair et al. 1999, Soranzo et al. 2000, Cheddadi et al. 2006). Recent work has shown the presence of a previously unknown glacial refugium in northern Europe during the last glaciation (Naydenov et al. 2007, Pyhäjärvi et al. 2008), and of several small putative refugial areas in Southern Europe that gave rise to geographically limited “interglacial refugia” (Cheddadi et al. 2006). Among the southernmost populations the Spanish, Balkan and Turkish ones have usually been considered in broad scale phylogeographic studies (Sinclair et al. 1999, Soranzo et al. 2000, Naydenov et al. 2007, Pyhäjärvi et al. 2008), whereas Alpine and Apennine populations have received less attention (Scalfi et al. 2009). In the few studies where they were extensively sampled it was demonstrated that Italian populations shared a common mitochondrial haplotype, and that they are different from the surrounding Austrian, Swiss, and French alpine populations (Cheddadi et al. 2006). Cheddadi et al. therefore hypothesised a common origin for Italian Scots pine populations from a refugial area in Southern Italy, even though the results by Puglisi and Attolico (2000), Labra et al. (2006) and Scalfi et al. (2009), obtained with more polymorphic genetic markers (respectively allozymes, ISSRs and SSRs), showed a marked differentiation between Alpine and Apennine populations, suggesting different recolonisation routes for the two mountain chains. The preservation of genetic resources of highly fragmented Apennine populations therefore appears to be a high priority challenge.

At the population level genetic structure and gene flow patterns have been studied in Spanish populations from the northern Meseta by Robledo-Arnuncio and Gil (2005) and Robledo-Arnuncio et al. (2005), and show high within-populations genetic diversity and extremely high levels of pollen flow over long distances (5% longer than 30 km). On the other hand, small relic Scots pine populations from the Apennines had significantly lower genetic diversity, and are differentiated from the alpine ones, possibly as a

consequence of progressive isolation since the early Holocene and their origin from different glacial refugia (Labra et al. 2006, Scalfi et al. 2009). Scalfi et al. (2009) however, hypothesised a possible different role of gene flow via-pollen and seed in determining this genetic differentiation.

In this study we surveyed the genetic variability of 21 Scots pine populations throughout the species distribution in the Italian peninsula, using 9 highly informative nuclear microsatellite (nSSR) markers. Our main aim was to assess the levels and distribution of genetic variability of this conifer by intensively sampling the entire Italian range of the species, and to investigate the presence of any cryptic genetic structure shaped by postglacial recolonisation which went undetected in previous studies based on organellar markers. We also discussed our results in the light of the European Directive 105/1999, with particular emphasis on the preservation and restoration of Scots pine genetic resources in the Alps and Apennines.

Material and Methods

Plant material

Twenty-one native populations of Scots pine were chosen within the natural range of diffusion of the species in Italy (Figure 1). Four of them (BOS, VEZ, PAS and CAS) are located in the hilly areas of Piedmont, two (TIC and OLG) grow in the upper part of Po valley and another (CRO) is found in the northern Apennines. All the others are distributed along the entire Italian Alpine region. Table 1 summarises names and locations of the populations analysed. Most of the populations belong to mixed forests, with the exception of populations CAR, FEN and SAV which are pure stands of Scots pine. In other locations accompanying species vary according to altitude and latitude: English oak, hornbeam and wild cherry (TIC), pubescent oak, flowering ash and juniper (CRO), sessile oak and European ash (OLG, MAS, GAR), sessile oak, black locust, elm and maples (BOS, VEZ, PAS, CAS), black pine (DOG), Norway spruce (VAL, SIU, BRU, COR, CLA), silver fir (TOC), larch (SAR) and mountain pine (BRI). Among the populations sampled, CRO, CAR, FEN, OLG, VAL, SIU and BRU are registered in the Italian National Book of Seed Stands for Scots pine, selected for their phenotypic characteristics and health status (Morandini and Magini 1975).

Twenty-four adult non-adjacent trees were chosen at random in each population. Since some individuals did not show reliable nSSRs banding patterns, they were excluded from the analysis. The number of these individuals varied among populations: from zero up to six. Consequently, the total number of

analysed individuals was 449. Needles collected from the trees were stored at –20°C until DNA extraction was carried out.

Molecular analysis

Frozen needles (100 mg of tissue) were powdered in liquid nitrogen and genomic DNA was extracted using the QIAGEN® DNeasy plant mini kit, according to the manufacturer's protocol. A total concentration of 20 ng/µl was obtained for each sample.

Twelve simple sequence repeat markers (SSR) were selected according to the literature, and tested on our plant material (Table 2). While the SPAC and SPAG series consisted of primers specific for *P. sylvestris* (Soranzo et al. 1998), the PtTX series included primers originally designed for *Pinus taeda*, but they also proved to be as useful as the markers selected for *P. sylvestris* (Elsik et al. 2000, Auckland et al. 2002, González-Martinez et al. 2004).

Polymerase Chain Reaction (PCR) amplifications were performed using a Perkin Elmer GeneAmp® PCR System 9600 thermal cycler. The protocol was slightly modified, according to the presence of a fluorochrome (IR-Dye 700 and IR-Dye 800) attached to each forward primer. Each amplification reaction contained 1X reaction buffer (Promega), 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.5 µM of each primer, 0.65 U of GoTaq DNA Polymerase (Promega), approximately 10 ng genomic DNA, and deionised water to a total volume of 13 µl.

The PCR profiles varied for different primers, and were adjusted for the presence of the forward labelled ones: for *P. sylvestris* specific primers the profiles included an initial step of 3 min at 94°C, followed by 5 cycles of touchdown consisting of 94°C for 30 s, 65°C for 30 s Δ↓ 1°C (SPAC 11.4 and SPAC 11.5) or 60°C for 30 s Δ↓ 1°C (SPAC 11.8, SPAC 12.5, SPAG 3.7 and SPAG 7.14), 72°C for 1 min, and subsequent 25 cycles of amplification consisting of 94°C for 30 s, 60°C for 30 s, 72°C for 1 min, and a final extension of 72°C for 10 min. Touchdown PCR was not necessary for SPAC 11.6, where the protocol consisted of a first step at 95°C for 5 min, and 35 cycles (denaturation at 94°C for 1.5 min, annealing at 55°C for 1.5 min and elongation at 72°C for 1.5 min), followed by 10 min at 72° C. For *P. taeda* primers PCR profiles consisted of an initial step of 5 min at 94°C, followed by 20 cycles of touchdown [94°C for 1 min, 59°C for 30 s Δ↓ 0,5°C (PtTX 3032) or 55°C for 30 s Δ↓ 0,5°C (PtTX 3116) or 60°C for 30 s Δ↓ 0,5°C (PtTX 3107, PtTX

4001 and PtTX 4011), 72°C for 1 min)], followed by 20 cycles of amplification at 94°C for 1 min, annealing temperature for 1 min, 72°C for 1 min, and a final extension of 72°C for 3 min.

The forward sequence of each primer pair was labelled with a fluorescent dye at its 5' end: IR-Dye 800 for SPAC 11.6, SPAC 12.5, SPAG 7.14, PtTX 3107 and PtTX 4001, IR-Dye 700 for SPAC 11.4, SPAC 11.8, SPAC 11.5, SPAG 3.7, PtTX 3032, PtTX 3116 and PtTX 4011.

Electrophoresis and detection of PCR products were carried on a 6%, 25 cm long, 0.25 mm thick, denaturing polyacryalmide gel using a sequencer (model DNA 4200 Sequencer LI-COR Biotechnology). Gels were run at 2000 V in TBE 1X buffer, for 1 to 3 h, depending on the product sizes. Determination of polymorphism was obtained using two different marked standards of known molecular weight (50–350 bp, and 50–700 bp). Data were collected by e-Seq software (DNA Sequencing and Analysis Software), and all the size scores were visually checked.

Data processing

Allele frequencies and within-population genetic diversity parameters (mean number of alleles per locus, *A*; mean number of private alleles per locus, P_a; effective number of alleles per locus, N_e; observed and expected heterozygosity, H_O and H_E, respectively) were estimated using GenAlEx v.6 software (Peakall and Smouse 2006). Allelic richness, based on a minimum sample size of 16 gene copies (A_{r16}), was calculated using FSTAT (Goudet 1995). Genotypic disequilibrium between pairs of loci was tested at the single population level and across all populations, with Fisher's exact test using Arlequin software (Excoffier et al. 2005).

Fisher's exact test using the Markov Chain algorithm (Guo and Thompson 1992) was used to assess deviations from Hardy-Weinberg equilibrium for each population and each locus, and where significant deficiencies of heterozygotes from Hardy-Weinberg expectations were found the presence of a relatively high frequency of null alleles was suspected (Pemberton et al. 1995). Loci with high frequencies of null alleles were identified by estimating null allele frequencies for each locus and each population, using the software Micro-Checker (Van Oosterout et al. 2004). In further analysis we eliminated problematical loci with high null allele frequency from the dataset, using only loci with < 0.19 null allele frequencies. This value has been considered as a threshold over which significant underestimate of H_E due to null alleles can

196 be found (Chapuis et al. 2008). Where possible, we employed robust analyses for the presence of null alleles
197 (Chapuis and Estoup 2007, Chapuis et al. 2008), including STRUCTURE (Pritchard et al. 2000, Falush et al.
198 2003), ordination methods (sPCA, Jombart et al. 2008), and the Mantel test on chord distance (DC, Cavalli-
199 Sforza and Edwards 1967). In particular the inbreeding coefficients (F_{IS}) were calculated taking into account
200 the estimated null allele frequencies using the program INEst, and running the individual inbreeding model
201 (IIM) with a Gibbs sampler of 10^5 iterations (Chybicki and Burczyk 2009).

202 FreeNA was used to compute the value of Weir's (1996) estimators of F-statistics to analyse the
203 population genetic structure of the overall samples. In particular, F_{ST} was calculated in order to estimate the
204 proportion of the total genetic variation due to differentiation among populations. Genetic differentiation
205 between populations was estimated computing a pairwise F_{ST}. FreeNA applies the ENA correction method to
206 efficiently correct for the positive bias induced by the presence of null alleles on the F_{ST} estimation (Chapuis
207 and Estoup 2007).

208 The genetic structure of the populations was explored using Bayesian clustering and spatial principal
209 components analysis (sPCA). Bayesian clustering was performed with the software STRUCTURE (Pritchard
210 et al. 2000). The program uses a Markov chain Monte Carlo (MCMC) algorithm to cluster individuals into
211 populations on the basis of multilocus genotypic data. Individual multilocus genotypes are first assigned
212 probabilistically to genetic clusters (K) without considering sampling origins. Admixed or hybrid individuals
213 can be identified as they will have a fraction of their alleles derived from each genetic cluster. The program
214 was run setting a burn-in period of 10^5 followed by $5x10^5$ iterations, and using the admixture ancestry model
215 and the correlated allele frequency model, given the low F_{ST} and the high genetic connectivity typical of
216 forest tree populations. Posterior probabilities of K were calculated from the means of 20 runs for each value
217 of $K \varepsilon \{1, ... 10\}$, and the optimum K determined using the method of Evanno et al. (2005).

218 The sPCA is a spatially explicit multivariate method recently developed by Jombart et al. (2008) to
219 investigate the spatial pattern of genetic variability using allelic frequency data of individuals or populations.
220 It takes spatial information directly into account as a component of the adjusted model, focusing on the part
221 of the variability that is spatially structured. This analysis does not require data to meet the Hardy–Weinberg
222 expectations, or linkage equilibrium to exist between loci. The sPCA yields scores summarising both the
223 genetic variability and the spatial structure among individuals (or populations). Global structures (patches,

clines and intermediates) are disentangled from local ones (strong genetic differences between neighbours) and from random noise. Neighbouring sites were defined by building a connection network based on Delaunay triangulation. The existence of global and local structuring was tested using the multivariate Monte Carlo tests implemented, as the sPCA procedure, in the *adegenet* package for R (Jombart 2008; R Development Core Team, 2009).

A Mantel (1967) test was applied to the matrices of pairwise chord distance and log-transformed geographical distance between populations (natural logarithm scale) to assess isolation-by-distance, namely the model under which genetic differentiation between populations is the result of drift. Chord distance (Cavalli-Sforza and Edwards 1967) for each pair of populations was calculated using the INA correction described in Chapuis and Estoup (2007) with FreeNA. The test of significance for Mantel test was carried out on 9,999 permutations of the data. Mantel test was executed on the entire dataset and on clusters detected by STRUCTURE and sPCA analyses.

Finally, we used the program Bottleneck v.1.2.02 (Piry et al. 1999) to test for recent population bottlenecks. A Wilcoxon's sign rank test was used to compare expected heterozygosity from Hardy-Weinberg equilibrium with predicted heterozygosity at mutation-drift equilibrium, on the basis of the observed allele number (Piry et al. 1999). The program was run under a two-phase model of mutation (TPM) that generally fits microsatellite evolution better than either pure stepwise or infinite allele models (Di Rienzo et al. 1994). One thousand simulations were performed for each sample based on a TPM consisting of 90% single-step mutations and 10% multistep changes.

Results

Allelic diversity of microsatellite loci

Ten out of the 12 tested microsatellites (SPAC 11.4, SPAC 11.6, SPAC 11.8, SPAC 12.5, SPAG 7.14, PtTX3032, PtTX3107, PtTX3116, PtTX4001 and PtTX4011) showed reliable banding patterns with clear and repeatable bands. On the contrary, SPAC 11.5 did not amplify in any sample, and SPAC 3.7 failed to amplify the DNA in the majority of samples. The latter two markers were therefore excluded from the analysis. SPAC 11.8 was also excluded due to the high frequency of possible null alleles (0.57). The

presence of null alleles was also suspected for 6 out of the remaining 9 loci, with lower frequencies ranging from 0.07 (PtTX3032) to 0.19 (SPAC 11.6).

The 9 selected microsatellites were highly polymorphic, and generated a total of 532 alleles (range 21 to 151), with a mean number of ~48 alleles per locus. It was also possible to detect 137 private alleles, i.e. present only in one population. The population frequency of private alleles was on average 0.036 and ranged between 0.021 and 0.208. The distribution of these alleles among loci ranged from 43 (PtTX3032) to 1 (PtTX3107). PAS (17 alleles) was the population with the highest number of private alleles, followed by TOC (14 alleles) and COR (10 alleles), and three populations (BOS, CLA and GAR) presented only two private alleles each.

Genetic variation within populations

Statistics on the genetic diversity within populations are given in Table 3. A high level of intra-population variability was found, since on average more than 14 alleles per locus were observed (A = 14.45). CRO (the only populations from the Apennines) showed the lowest value of genetic diversity (A = 10.4, N_e = 5.4, A_{r16} = 6.781, and H_e = 0.754). The populations belonging to the Po valley (TIC and OLG) also showed a low variability, as well as the populations from the hills of Piedmont (BOS, VEZ, PAS, and CAS), even though the genetic impoverishment was less pronounced. Except for these particular populations (which, in most cases, are small, isolated, and/or on the edge of the main range) only BRI, GAR, and CLA showed a more or less marked reduction in genetic diversity. This pattern was confirmed for all genetic diversity parameters considered.

Despite the fact that some of the investigated populations are characterised by a low genetic diversity, no evidence for recent bottlenecks was found. In fact, for all the populations HW heterozygosity and expected gene diversity at mutation drift equilibrium did not differ significantly.

The probability that two randomly sampled alleles in a given population were not the same was higher than 84% (mean H_e = 0.847), whereas the observed heterozygosity (mean H_o = 0.670) was lower than expected. The difference, that determines a significant positive value for mean inbreeding coefficient, is mainly due to non-random mating and null alleles. Recalculating the inbreeding coefficients, taking into account the frequencies of null alleles, we found that deviations from the Hardy-Weinberg equilibrium were

low (F_{IS} ranging from 0.052 in FEN to 0.016 in TIC, with a mean of 0.033). None of estimated F_{IS} was significantly different from zero (confidence interval calculated through INEst estimates of F_{IS} overlapped zero in all populations).

Genetic differentiation among populations

Most of the genetic diversity was found within populations, while a small amount of the variability occurred among populations (F_{ST} = 0.058, CI: 0.037-0.081). The F_{ST} values per locus ranged from 0.024 (SPAC 11.6) to 0.12 (PtTX 4011), and there were no obvious differences between the *P. sylvestris* and the *P. taeda* sets of markers. The genetic divergence between populations was further investigated by computing a pairwise F_{ST} matrix. Multilocus F_{ST} values varied between 0.015 (CAR and SAR) and 0.141 (CRO and VEZ). The population from the Apennines (CRO) was always clearly separated by the others. Almost all pairwise F_{ST} values were significantly greater than zero, confirming the presence of a slight, although significant, amount of population structuring in Italian Scots pine (results not shown).

Following the method of Evanno et al. (2005), the Bayesian clustering results obtained with STRUCTURE indicate that K=2 clusters represents the most informative representation of the overall genetic structure that we analysed (Figure 2). We found that most individuals from western populations (VEZ, PAS, CAS, SAV, and TOC) clearly belong to cluster 2, whereas eastern populations and CRO (the Apennine population) are primarily composed by individuals from cluster 1, with the exception of some admixed populations (COR, SIU, and VAL). Although populations belonging to the pedo-climatic region of the Po valley (TIC and OLG) are closer to western populations, their individuals are predominantly assigned to cluster 1. This East-West subdivision was confirmed by sPCA analysis (Figure 3). The existence of such 'global structure' (*sensu* Jombart et al., 2008) was demonstrated using both Delaunay triangulation (tmax = 0.0631, $P < 0.05$) and Gabriel graph (tmax = 0.0672, $P < 0.05$) for building the connection network.

The correlation between genetic diversity, expressed as Cavalli-Sforza and Edwards (1967) chord distance for pairs of populations, and the logarithm of distances expressed in km, did not show the typical pattern of isolation by distance, and did not suggest a strong relationship between the two factors. Mantel's test was not significant when performed on the entire dataset ($P = 0.094$), as well as when performed

306 separately on the two clusters detected by sPCA analysis (P = 0.669 in eastern populations, and P = 0.482 in
307 western populations excluding CRO).
308
309 **Discussion and Conclusion**
310 The aim of this study was to assess the level and the distribution of genetic variation of Scots pine
311 throughout its natural range in Italy, in order to get fundamental knowledge that can be applied for plant
312 propagation and genetic resources conservation. The results could be slightly biased due to the limited
313 number of individuals sampled per populations. Although a larger database (more than 50 individuals per
314 population) would be preferred to obtain stronger data on genetic variation at the level of polymorphic loci
315 (Nei 1978), phylogeographic studies using nSSR markers often are based on less than 30 individuals per
316 populations (e.g. Williams et al. 2007, Ferrazzini et al. 2008, Bagnoli et al. 2009, Scalfi et al. 2009, Bai et al.
317 2010). Kalinowski (2005) demonstrated that some genetic distances, such as F_{ST}, showed limited sampling
318 variance at loci characterised by a high mutation. He also showed that increasing the number of loci, instead
319 of increasing the number of individuals, is an effective way to improve the precision of measures of genetic
320 differentiation. Miyamoto et al. (2008) showed by resampling simulations, that an accurate estimate of
321 genetic diversity (H_e) can also be achieved with small samples (less than 30 plants per population) genotyped
322 at nSSRs. On the other hand, larger sample sizes are needed in order to obtain more accurate estimates of
323 allelic richness, although the statistical technique of rarefaction can compensate for sampling disparities and
324 allow for meaningful comparisons among populations. In general, in the present paper we used population
325 genetic indexes and techniques that minimize possible estimation biases caused by small sample size.
326 The populations analysed showed a considerable amount of genetic diversity, as estimated by means
327 of variation scored at nine nuclear microsatellite loci. The observed number of alleles per locus (A) ranged
328 from 10.4 to 18.4, with an average per population of 14.45, and the average gene diversity (H_e) was as high
329 as 0.847. The high degree of observed diversity is not surprising since it has been recognised for a long time
330 as a peculiar characteristic of woody plants (Hamrick et al. 1992). Furthermore, species such as Scots pine,
331 which do not have a strong habitat specificity, and are almost continuously distributed, are expected to have
332 more within-population diversity than those with strong habitat preference and a scattered distribution. Data
333 from this analysis are coherent with those reported in the literature: for example Robledo-Arnuncio et al.

(2005) estimated nSSR loci values of 23.0 and 0.923 for A and H_e respectively. However, this Authors used only three loci, but if we reference our data with these markers we obtained values of 19.4 (A) and 0.885 (H_e).

The Apennine population (CRO) showed lower values for all the calculated genetic variation indices. This confirms the genetic erosion undergone by this population, most likely as a consequence of isolation and limited population size. Lower values of genetic diversity in Apennine populations, compared to Alpine ones, were already recorded by Scalfi et al. (2009) at nSSR loci, and by Puglisi and Attolico (2000) and Labra et al. (2006), using different genetic markers, respectively isozymes and ISSRs. Similar patterns were also observed with reference to populations from the Po valley (TIC and OLG). As for CRO, habitat fragmentation could be the main cause of the genetic diversity reduction detected for these populations. TIC and, to a lesser extent, OLG can be considered as relic populations, especially after the reduction of the Scots pine range in the Po valley due to intensive use of land for both agriculture and urbanisation purposes. These non-alpine populations are nowadays regressing, mainly due to problems of seed dispersal, competition with other forest trees and as a consequence of rural depopulation and abandonment of wood management (Camerano 2008, Regione Piemonte and Regione Valle d'Aosta 2008). Their preservation is therefore a primary goal: it is needed to maintain genetic diversity as well as the forest landscape. Particular attention should be addressed to the population PAS. It is a very small population, where only a few dozens of individuals still survive, threatened by the competition with other species (namely black locust). This population showed the highest number of private alleles (17) confirming an ongoing genetic erosion process which could cause the local extinction of the population over a short period of time.

The overall level of genetic diversity arising from the differentiation between populations found in this study (F_{ST} = 0.058) is moderate, but higher than that previously observed in the same species in other European countries (Müller-Starck et al. 1992). In the Scandinavian region, for instance, values of $F_{ST} \leq 0.02$ were found in populations of Scots pine studied with different markers (Karhu et al. 1996). Allozymes gave a F_{ST} = 0.03 between Sweden and Siberian populations (Wang et al. 1991), and a G_{ST} = 0.021 ÷ 0.046 among many European stands (Prus-Glowacki et al. 1993, Prus-Glowacki et al. 2003). The F_{ST} value found in our study is interesting, especially if we consider the relatively small geographic distances between the Italian

362 populations. A low level of genetic differentiation among populations is common in conifers. They maintain
363 most of their variation within populations (Hamrick et al. 1992), which can be explained by the mainly
364 allogamous mating system, and by the high gene flow rate favoured by their dispersal strategy and
365 widespread diffusion (Petit and Hampe 2006, Piotti et al. 2009, Williams 2010). Mantel's test, performed in
366 order to check the presence of IBD between the populations studied was not significant, and thus IBD was
367 apparently not a mechanism shaping the present distribution of genetic variability.

368 The analysis of genetic differentiation, and the most likely population clustering according to
369 STRUCTURE and sPCA analyses indicate that the global structure detected separate gene pools for the
370 eastern and the western Alps. Despite the higher similarity of the Apennine population with the eastern Alps
371 populations, rather than with closer populations from the western Alps, we found a generally high
372 differentiation between populations from the two mountain chains. These two results seem to exclude a
373 common postglacial origin of Italian *P. sylvestris* populations based on the shared RFLP and *nad* 1 intron
374 mitotypes (Sinclair et al. 1999, Cheddadi et al. 2006, Labra et al. 2006). Our results support the hypotheses
375 of an Apennine glacial refugium (see for instance Puglisi and Attolico 2000), but with no evidence of any
376 expansion from Apennines into the southern slope of the Alps, as hypothesised by Cheddadi et al. (2006).
377 The *P. sylvestris* Apennine population (CRO) is genetically distinct from those in the western Alps, although
378 a certain level of admixture exists within the BOS and CAR populations. Similar results have been recently
379 obtained by Piovani et al. (2010), studying *Abies alba* populations. These authors also found that Apennine
380 populations are genetically different from populations from the western Alps. Recent studies based on the
381 analysis of stratigraphic records of pollen, stomata, and macrofossils in northern Italy showed that, although
382 the southern slope of the Alps was extensively glaciated during the last glacial maximum (LGM), conifer and
383 several broad-leaved tree species survived in the Po plain and along the southeastern Alpine border (Vescovi
384 et al. 2007). It has also been recently demonstrated that *P. sylvestris* survived the LGM in the Euganean Hills
385 (north-eastern Italy), a hilly area 50-60 km south of the maximum extent of the last Alpine glaciation
386 (Kaltenrieder et al. 2009). An early presence (ca 15000 BP) of *P. sylvestris* after LGM was also signalled by
387 Finsinger et al. (2006) in the western Alps, at Lago piccolo di Avigliana (Piedmont). Our results depict a
388 scenario where populations from the western Alps, eastern Alps and the Apennines originated from at least
389 three different refugia, with possible contact zones between the western and eastern Alps,

colonisation/expansion routes at the latitude of the TIC and TOC populations, and between the western Alps and the Apennine populations in southern Piedmont (BOS and CAR populations). In addition, our study presents further evidence of the possible past genetic connection of western Alps and Apennine populations through the Po plain, a vast area (ca 200 km large) where *P. sylvestris* was widespread during the early Holocene (Labra et al. 2006, Scalfi et al. 2009).

The management of genetic resources may benefit from the identification of genetically homogeneous regions, since genetic pollution of local genetic stocks by plantations of non-local origin material can be greatly reduced by the use and transfer of suitable propagation material. Assuming local populations to be optimal as basic material for local uses, genetic zones would have relevance for defining variation in potential commercial and adaptive traits; and they would provide the genetic background required in order to establish the number and location of primary gene-pool reserves for the species. Moreover, the definition of genetic parameters based on molecular markers for each genetic zone can be useful for certification of seedlots and breeding programmes (Bucci and Vendramin 2000). Since 1999, the European Council Directive 1999/105/CE regulates the forest reproductive material market and transfer in Europe. The Italian Government has implemented this directive, with the Decree No. 386/2003. One of the most important features of the acts is the definition of region of provenance as "the area or group of areas subjected to sufficiently uniform ecological conditions in which stands or seed sources showing similar phenotypic or genetic characters are found". The identification of these areas plays a basic role for a rational management of activities linked with forest tree propagation, including afforestation and in situ genetic preservation.

The results of our study contribute to a better understanding of our knowledge on genetic variation of Scots pine in Italy. The information is of basic importance for the definition of regions of provenance for Scots pine, although a deeper knowledge of the ecological characteristics of the areas of the study, such as vegetational and phytogeographical data, is also needed.

Acknowledgments

The research was supported by the project "RI.SELV.ITALIA" (task 1.1, Biodiversity and production of forest reproductive material) from the Italian Ministry for Agriculture, Food and Forestry

418 Policy. Thanks are also due to the Italian Ministry of Education, Universities and Research, which partially
419 granted the work (local research funds). Thanks are also due to colleagues from “Translations Group” for the
420 language revision of the manuscript (www.translation-group.com/).

421

422 **References**

423

424 Auckland LD, Bui T, Zhou Y, Shepherd M, Williams CG (2002) Conifer microsatellite handbook. Corporate
425 Press, Raleigh, NC, USA
426
427 Bagnoli F, Vendramin GG, Buonamici A, Doulis AG, González-Martínez SC, La Porta N, Magri D, Raddi,
428 P, Sebastiani F, Fineschi S (2009) Is *Cupressus sempervirens* native in Italy? An answer from genetic and
429 palaeobotanical data. Mol Ecol 18:2276-2286
430
431 Bai WN, Liao WJ, Zhang DY (2010) Nuclear and chloroplast DNA phylogeography reveal two refuge areas
432 with asymmetrical gene flow in a temperate walnut tree from East Asia. New Phytol 188:892-901
433
434 Bucci G, Vendramin GG (2000) Delineation of genetic zones in the European Norway spruce natural range:
435 preliminary evidence. Mol Ecol 9:923-934
436
437 Camerano P, Gottero F, Terzuolo P, Varese P (2008) Tipi forestali del Piemonte, 2nd edition. Regione
438 Piemonte, Blu Edizioni, Torino
439
440 Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. Am J
441 Hum Genet 19:233– 257
442
443 Chapuis M, Estoup A. (2007) Microsatellite Null Alleles and Estimation of Population Differentiation. Mol
444 Biol Evol 24:621-631
445
446 Chapuis M, Lecoq M, Michalakis Y, Loiseau A, Sword GA, Piry S, Estoup A (2008) Do outbreaks affect
447 genetic population structure? A worldwide survey in *Locusta migratoria*, a pest plagued by microsatellite
448 null alleles. Mol Ecol 17: 640-3653
449
450 Cheddadi R, Vendramin GG, Litt T, François L, Kageyama M, Lorentz S, Laurent JM, De Beaulieu JL,
451 Sadori L, Jost A, Lunt D (2006) Imprints of glacial refugia in the modern genetic diversity of *Pinus*
452 *sylvestris*. Global Ecol Biogeogr 15:271-282
453
454 Chybicki I, Burczyk J (2009) Simultaneous estimation of null alleles and inbreeding coefficients. J Hered
455 100:106-113
456
457 Di Rienzo AA, Peterson C, Garza JC, Valdes AM, Slatkin M, Freime RB (1994) Mutational processes of
458 simple-sequence repeat loci in human populations. P Natl Acad Sci USA 91:3166-3170
459
460 Eckert CG, Samis K, Lougheed SC (2008) Genetic variation across species' ranges: the central-marginal
461 hypothesis and beyond. Mol Ecol 17:1170-1188
462
463 Elsik CG, Minihan VT, Hall SE, Scarpa AM, Williams CG (2000) Low-copy microsatellite markers for
464 *Pinus taeda* L. Genome 43:550-555
465
466 Elsik CG, Williams CG (2001) Low-copy microsatellite recovery from a conifer genome. Theor Appl Genet
467 103:1189-1195
468

Escudero A, Iriondo JM, Torres ME (2003) Spatial analysis of genetic diversity as a tool for plant conservation. Biol Conserv 113:351–365

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

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evol Bioinform 1:47-50

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567-87

Ferrazzini D, Monteleone I, Belletti P (2008) Small-scale genetic diversity in oneseed hawthorn (*Crataegus monogyna* Jacq.). Eur J For Res 127:407-414

Finsinger W, Tinner W, Vanderknaap W, Ammann B (2006) The expansion of hazel (*Corylus avellana* L.) in the southern Alps: a key for understanding its early Holocene history in Europe? Quaternary Sci Rev 25:612-631

González-Martinez SC, Robledo-Arnuncio JJ, Collada C, Díaz A, Williams CG, Alía R, Cervera MT (2004) Cross-amplification and sequence variation of microsatellite loci in Eurasian hard pines. Theor Appl Genet 109:103-111

Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate *F*-statistics. J Hered 86:485-486

Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361-372

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. New Forest 6:95-124

Jombart T (2008) *adegent*: a R package for the multivariate analysis of genetic markers. Bioinformatics 11:1403-1405

Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. Heredity 101:92-103

Jump AS, Peñuelas J (2006) Genetic effects of chronic habitat fragmentation in a wind-pollinated tree. P Natl Acad Sci USA 103:8096-8100

Kalinowski ST (2005) Do polymorphic loci require large sample sizes to estimate genetic distances? Heredity 94:33-6

Kaltenrieder P, Belis CA, Hofstetter S, Ammann B, Ravazzi C, Tinner W (2009) Environmental and climatic conditions at a potential Glacial refugial site of tree species near the Southern Alpine glaciers. New insights from multiproxy sedimentary studies at Lago della Costa (Euganean Hills, Northeastern Italy). Quaternary Sci Rev 28:2647-2662

Karhu A, Hurme P, Karjalainen M, Karvonen P, Kärkkäinen K., Neale D, Savolainen O (1996) Do molecular markers reflect patterns of differentiation in adaptive traits of conifers? Theor Appl Genet 93:215-221

Kramer AT, Ison JL, Ashley MV, Howe HF (2008) The Paradox of Forest Fragmentation Genetics. Conserv Biol 4:878-885

Labra M, Grassi F, Sgorbati S, Ferrari C (2006) Distribution of genetic variability in southern populations of Scots pine (*Pinus sylvestris* L.) from the Alps to the Apennines. Flora 201:468-476

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209-220

Miyamoto N, Fernandez-Manjarres JF, Morand-Prieur ME, Bertolino P, Frascaria-Lacoste N (2008) What sampling is needed for treliable estimations of genetic diversity in *Fraxinus excelsior* L. (Oleaceae)? Ann Sci 65: 403-410.

Morandini R, Magini E (1975) Il materiale forestale di propagazione in Italia. Collana Verde n. 34, MAF – Corpo Forestale dello Stato, Roma

Müller-Starck G, Baradat PH, Bergmann F (1992) Genetic variation in European tree species. New Forests 6: 23-47

Nathan R (2006) Long-distance dispersal of plants. Science 313:786-788

Naydenov K, Senneville S, Beaulieu J, Tremblay F, Bousquet J (2007) Glacial vicariance in Eurasia: mitochondrial DNA evidence from Scots pine for a complex heritage involving genetically distinct refugia at mid-northern latitudes and in Asia Minor. BMC Evol Biol 7:233

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583-590

Palmberg-Lerche C (2001) Conservation of forest biological diversity and forest genetic resources. In: Palmberg-Lerche C, Iversen PA, Sigaud P (eds), Forest Genetic Resources Bulletin 29:24-26

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

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288-295

Pemberton JM, Slate J, Bancroft DR, Barrett JA (1995) Nonamplifying alleles at microsatellites loci: a caution for parentage and population studies. Mol Ecol 4:249-252

Petit R, Aguinagalde, I, Beaulieu, J-L, Bittkau, C., Brewer, S, Cheddiari, R, Ennos, R, Fineschi, S, Grivet, D, Lascoux, M, Mohany, A, Müller-Starck, G, Demesure-Musch, B, Palme, A, Martin, J, Rendell, S, Vendramin, G (2003): Glacial refugia: hotspots but not melting pots of genetic diversity. Science 300: 1563-1565

Petit R, Hampe A (2006) Some evolutionary consequences of being a tree. Annu Rev Ecol Evol S 37;187-214

Pignatti A (1982) Flora d'Italia. Vol. III. Edagricole, Bologna, Italy

Piotti A, Leonardi S, Piovani S, Scalfi M., Menozzi P (2009) Spruce colonization at treeline: where do those seeds come from? Heredity 103:136-145

Piovani P., Leonardi S, Piotti A, Menozzi P. Conservation genetics of small relic populations of Silver Fir (*Abies alba* Mill.) in northern Apennines. Plant Biosyst 144:683-691.

Piry S, Luikart G, Cornuet J. (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. J Hered 90:502–503

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

Prus-Glowacki W, Urbaniak L, Zubrowska-Gil M (1993) Allozyme differentiation in some European populations of Scots pine (*Pinus sylvestris* L.). Genetica Polonica 2:159-176

Prus-Głowacki W, Stephan BR, Bujas E, Alia R, Marciniak A (2003) Genetic differentiation of autochthonous populations of *Pinus sylvestris* (Pinaceae) from Iberian peninsula. Plant Syst Evol 239:55–66

Puglisi S, Attolico M (2000) Allozyme variation in natural populations of the Italian range of *Pinus sylvestris* L. For Genet 7**:**221–232

Pyhäjärvi T, Salmela MJ, Savolainen O (2008) Colonization routes of *Pinus sylvestris* inferred from distribution of mitochondrial DNA variation. Tree Genet Genomes 4:247-254

R Development Core Team (2009) R: A language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Regione Piemonte, Regione Autonoma Valle d'Aosta (2008) Il deperimento del pino silvestre nelle Alpi occidentali: natura e indirizzi di gestione. Compagnia delle Foreste, Arezzo.

Reich PB, Oleksyn J (2008) Climate warming will reduce growth and survival of Scots pine except in the far north. Ecol Lett 11:588-597

Robledo-Arnuncio JJ, Collada C, Alia R, Gil L (2005) Genetic structure of mountain isolates of *Pinus sylvestris* L. in a Mediterranean refugial area. J Biogeogr 32:595-605

Robledo-Arnuncio JJ, Gil L (2005) Patterns of pollen dispersal in a small population of *Pinus sylvestris* L. revealed by total-exclusion paternity analysis. Heredity 94:13-22

Rubiales MJ, Garcia-Amorena I, Alvarez SG, Manzaneque FG (2008) The Late Holocene extinction of *Pinus sylvestris* in the western Cantabrian Range (Spain). J Biogeogr 10:1840-1850

Scalfi M, Piotti A, Rossi M, Piovani P (2009) Genetic variability of Italian southern Scots pine (*Pinus sylvestris* L.) populations: the rear edge of the range. Eur J Forest Res 128:377-386

Sinclair WT, Morman JD, Ennos R (1999) The postglacial history of Scots pine (*Pinus sylv*estris L.) in Western Europe: evidence from mitochondrial DNA variation. Mol Ecol 8:83-88

Soranzo N., Alia R, Provan J, Powell W (2000) Patterns of variation at a mitochondrial sequence-tagged-site locus provides new insights into the postglacial history of European *Pinus sylvestris* populations. Mol Ecol 9:1205-1211

Soranzo N, Provan J, Powell W (1998) Characterization of microsatellite loci in *Pinus sylvestris* L. Mol Ecol 7:1260-1261

Toro MA, Caballero A (2005) Characterization and conservation of genetic diversity in subdivided populations. Phil T Roy Soc B 1459:1367-1378

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 40:535-538

Vescovi E, Ravazzi C, Arpenti E, Finsinger W, Pini R, Valsecchi V, Wick L, Ammann B, Tinner W (2007) Interactions between climate and vegetation during the Late glacial period as recorded by lake and mire sediment archives in Northern Italy and Southern Switzerland. Quaternary Sci Rev 26:1650–1669

636 Wang XR, Szmidt AE, Lindgren D (1991) Allozyme differentiation among populations of *Pinus sylvestris* L.
637 from Sweden and China. Hereditas 114:219-226
638
639 Weir BS (1996) Genetic Analysis Data II. Sinauer Associates, Sunderland, Mass
640
641 Williams CG (2010) Long-distance pine pollen still germinates after meso-scale dispersal. Am J Bot 97:1-10
642
643 Williams DA, Wang YQ, Borchetta M, Gaines MS (2007) Genetic diversity and spatial structure of a
644 keystone species in fragmented pine rockland habitat. Biol Conserv 138:256-268
645
646 Zhou Y, Bui T, Auckland LD, Williams CG (2002) Undermethylated DNA as a source of microsatellites
647 from a conifer genome. Genome 45:91-99

FIGURE LEGENDS

Figure 1. Geographical distribution of the 21 populations of Scots pine analysed in this study.

Figure 2. STRUCTURE results. On the top the relationship between K (number of inferred clusters) and Ln(K) and DeltaK, respectively. On the bottom, the probability of belonging to each of the two inferred clusters, according to the method by Evanno et al. (2005) for each of the individuals.

Figure 3. On the left the geographic representation of the connection matrix based on the Delaunay triangulation. On the right the geographic distribution of the first positive sPCA scores.

TABLE TITLES

Table 1. Details of site characteristics of Scots pine populations from Italy which were sampled for the study. Data on annual and summer rainfall, and average annual temperature were inferred for a period of at least five years.

Table 2. Descriptive statistics for the twelve microsatellite loci considered for the study.

Table 3. Statistics of genetic variation within Scots pine populations at nine microsatellite loci. A, mean number of alleles per locus; N_e, effective number of alleles per locus; A_{r16} Allelic richness based on a minimum sample size of 16 gene copies; P_a mean number of private alleles per locus; H_O, average observed heterozygosity; H_E, average gene diversity or expected heterozygosity; F_{IS}, average inbreeding coefficient, calculated taking into account the estimated null allele frequencies. Values in parenthesis are standard errors.

671 Table 1

Code	Population	Province	Location	Average elevation (m a.s.l.)	Annual (and summer) rainfall (mm)
CRO	Vezzano sul Crostolo	Reggio Emilia	44° 31’ N, 10° 31’ E	450	945 (174)
CAR	Carpe	Imperia	43° 55’ N, 7° 47’ E	1000	1100 (260)
BRI	Carnino Briga Alta	Cuneo	44° 18’ N, 7° 43’ E	1200	1170 (378)
BOS	Bossolasco	Cuneo	44° 32’ N, 8° 03’ E	750	877 (145)
VEZ	Vezza d’Alba	Cuneo	44° 45’ N, 8° 00’ E	300	660 (125)
FEN	Fenestrelle	Torino	45° 02’ N, 7° 03’ E	1450	900 (300)
SAV	Savoulx	Torino	45° 05’ N, 6° 40’ E	1300	396 (102)
PAS	Passerano Marmorito	Asti	45° 05’ N, 8° 05’ E	225	554 (156)
CAS	Casalborgone	Torino	45° 10’ N, 8° 00’ E	500	673 (142)
SAR	Sarre	Aosta	45° 43’ N, 7° 15’ E	1300	698 (174)
TOC	Toceno	Verbania	46° 10’ N, 8° 32’ E	1100	698 (459)
TIC	Ticino	Novara	45° 30’ N, 8° 39’ E	250	1020 (234)
OLG	Olgelasca	Como	45° 44’ N, 9° 11’ E	350	1635 (587)
MAS	Val Masino	Sondrio	46° 09’ N, 9° 34’ E	300	458 (165)
GAR	Valvestino Garda	Brescia	45° 45’ N, 10° 35’ E	800	685 (263)
VAL	Valda	Trento	46° 12’ N, 11° 16’ E	950	750 (350)
SIU	Alpe di Siusi	Bolzano	46° 33’ N, 11° 33’ E	1200	800 (400)
BRU	Brunico	Bolzano	46° 48’ N, 11° 56’ E	1100	692 (360)
COR	Cortina	Belluno	46° 32’ N, 12° 18’ E	1200	1100 (383)
CLA	Claut	Pordenone	46° 16’ N, 12° 31’ E	600	1186 (436)
DOG	Val Dogna	Udine	46° 26’ N, 13° 19’ E	1000	908 (365)

672

673 Table 2
674

Locus	Repeat motif	Primer sequences (5' → 3')	Number of alleles	Molecular weight range (bp)
SPAC 11.4[a]	$(AT)_5(GT)_{19}$	TCACAAAACACGTGATTCACA GAAAATAGCCCTGTGTGAGACA	38	130-170
SPAC 11.5[a]	$(AT)_8(GT)_{19}$-$(TA)_{11}$	TGGAGTGGAAGTTTGAGAAGC TTGGGTTACGATACAGACGATG	no amplification	
SPAC 11.6[a]	$(CA)_{29}(TA)_7$	CTTCACAGGACTGATGTTCA TTACAGCGGTTGGTAAATG	76	103-220
SPAC 11.8[a]	$(TG)_{16}$	AGGGAGATCAATAGATCATGG CAGCCAAGACATCAAAAATG	25	123-181
SPAC 12.5[a]	$(GT)_{20}(GA)_{10}$	CTTCTTCACTAGTTTCCTTTGG TTGGTTATAGGCATAGATTGC	62	116-202
SPAG 3.7[a]	$(TC)_{45}$	GTTAAAGAAAATAATGACGTCTC AATACATTTACCTAGAATACGTCA	no scorable bands	
SPAG 7.14[a]	$(TG)_{17}(AG)_{21}$	TTCGTAGGACTAAAAATGTGTG CAAAGTGGATTTTGACCG	64	174-252
PtTX 3032[b]	$(GAT)_{35}(GAC)_3GAT(GAC)_8$ $-(GAC)_6AAT(GAT)_6$	CTGCCACACTACCAACC AACATTAAGATCTCATTTCAA	151	254-572
PtTX 3107[c]	$(CAT)_{14}$	AAACAAGCCCACATCGTCAATC TCCCCTGGATCTGAGGA	22	144-175
PtTX 3116[c]	$(TTG)_7$-$(TTG)_5$	CCTCCCAAAGCCTAAAGAAT CATACAAGGCCTTATCTTACAGAA	67	100-276
PtTX 4001[d]	$(CA)_{15}$	CTATTTGAGTTAAGAAGGGAGTC CTGTGGGTAGCATCATC	31	197-231
PtTX 4011[d]	$(CA)_{20}$	GGTAACATTGGGAAAACACTCA TTAACCATCTATGCCAATCACTT	21	230-284

675

676 [a] Soranzo et al. 1998
677 [b] Elsik et al. 2000
678 [c] Elsik and Williams 2001
679 [d] Zhou et al. 2002

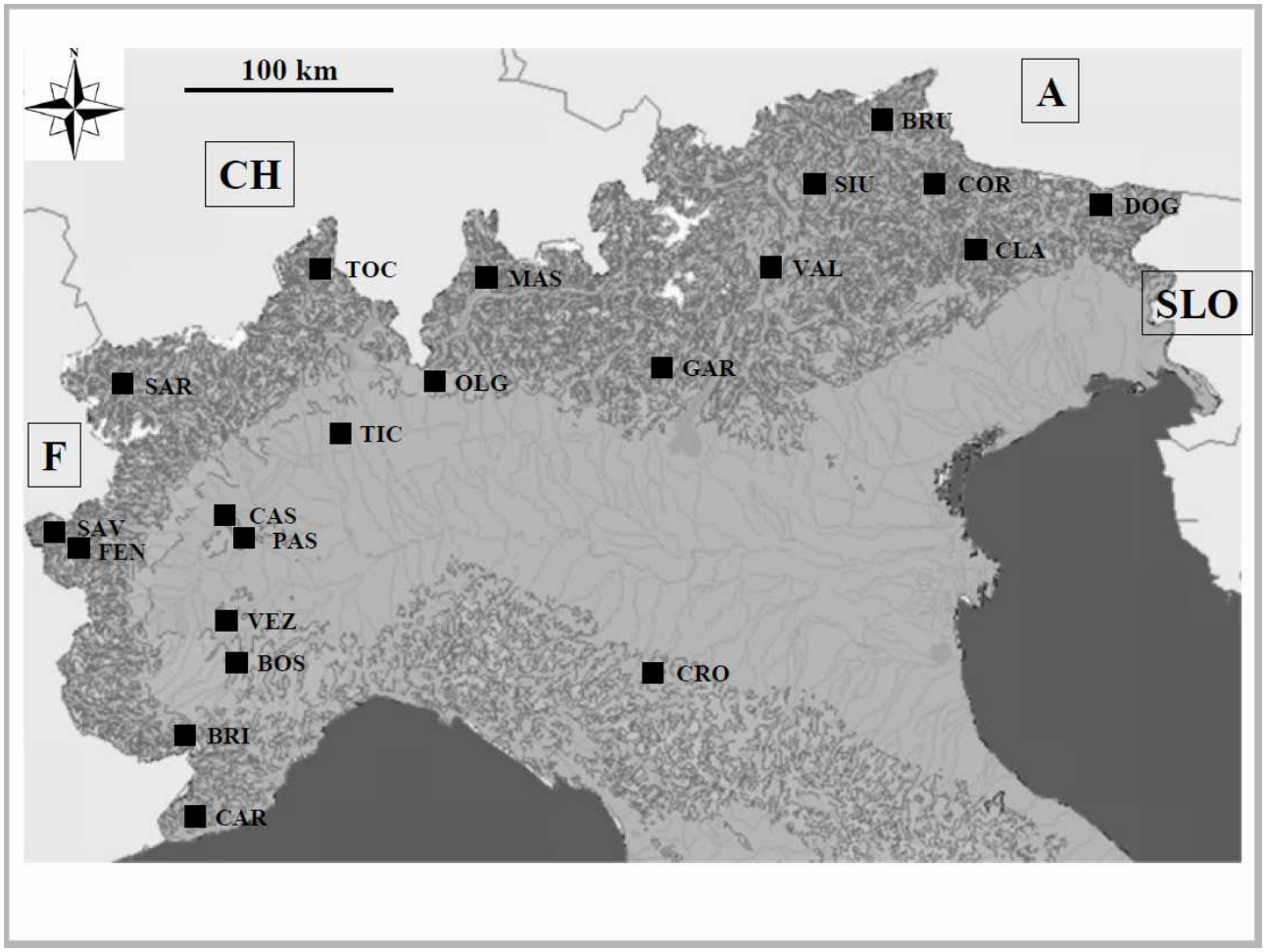
680 Table 3
681

Population	**A**	**N_e**	**A_{r16}**	**P_a**	**H_o**	**H_e**	**F_{IS}**
CRO	10.4	5.4	6.781	0.333	0.603	0.754	0.037
CAR	15.9	10.4	9.525	1	0.714	0.886	0.030
BRI	12.3	9.3	8.636	0.333	0.651	0.846	0.036
BOS	12.4	8.0	8.353	0.222	0.724	0.829	0.022
VEZ	14.6	9.4	8.939	0.778	0.670	0.852	0.032
FEN	16.0	10.2	9.391	0.556	0.664	0.845	0.052
SAV	18.4	12.2	10.003	1	0.665	0.877	0.023
PAS	14.1	10.1	9.830	1.889	0.576	0.874	0.049
CAS	13.9	8.7	8.601	0.778	0.645	0.858	0.031
SAR	16.4	10.9	9.641	0.889	0.698	0.876	0.029
TOC	17.6	11.0	9.637	1.556	0.641	0.850	0.025
TIC	12.9	6.9	7.713	0.778	0.703	0.822	0.016
OLG	11.2	7.2	8.049	0.444	0.621	0.811	0.047
MAS	16.3	10.6	9.363	0.889	0.703	0.855	0.049
GAR	13.1	8.2	8.398	0.222	0.661	0.808	0.028
VAL	15.3	10.6	9.793	0.333	0.772	0.872	0.022
SIU	15.7	9.7	9.137	0.889	0.676	0.854	0.028
BRU	16.0	9.8	9.326	0.444	0.630	0.865	0.021
COR	14.0	10.0	9.635	1.111	0.724	0.872	0.024
CLA	12.8	8.1	8.771	0.222	0.684	0.828	0.044
DOG	14.0	9.8	9.304	0.556	0.642	0.863	0.042
Overall mean	**14.45 (0.46)**	**9.37 (0.41)**	**8.992**	**0.724**	**0.670 (0.012)**	**0.847 (0.007)**	**0.033 (0.001)**

682

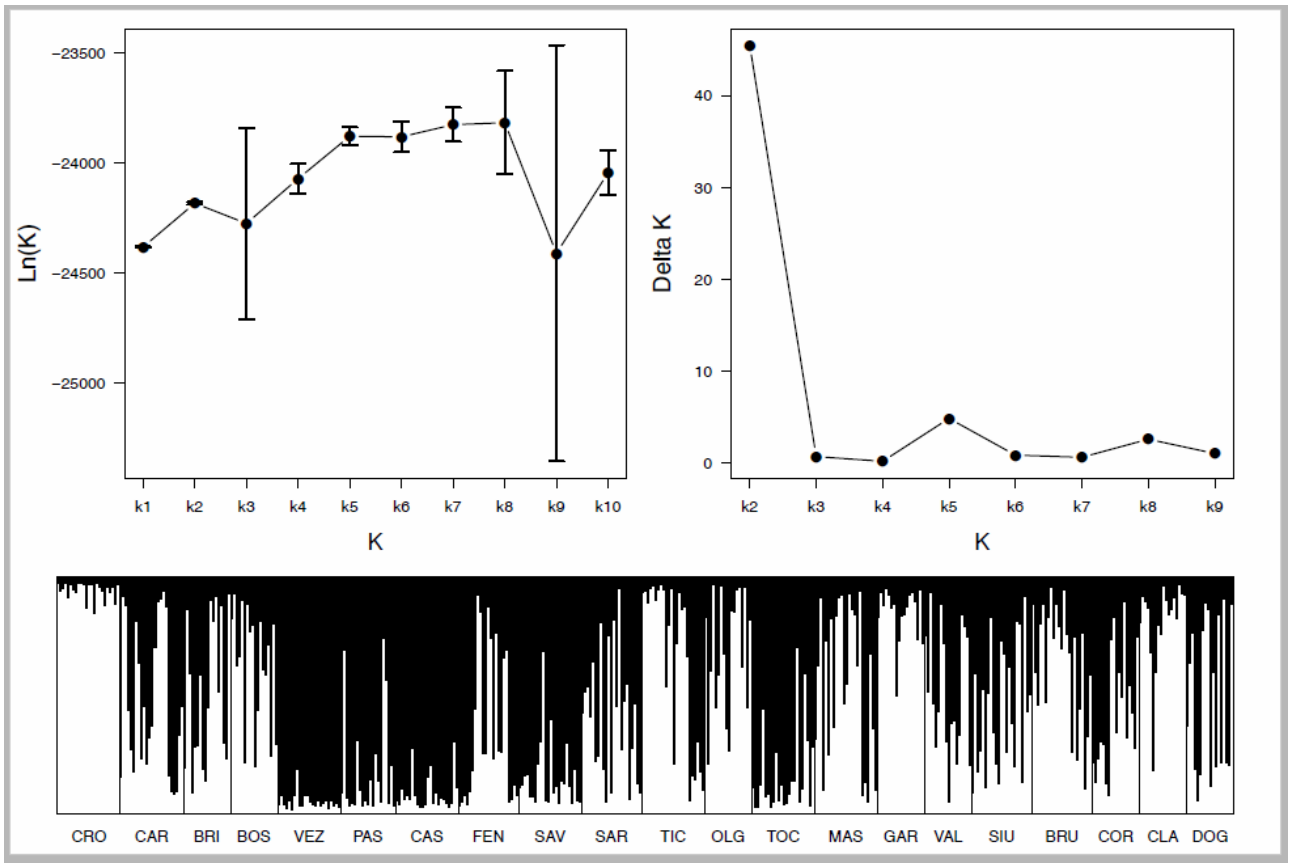
683 Figure 1

684



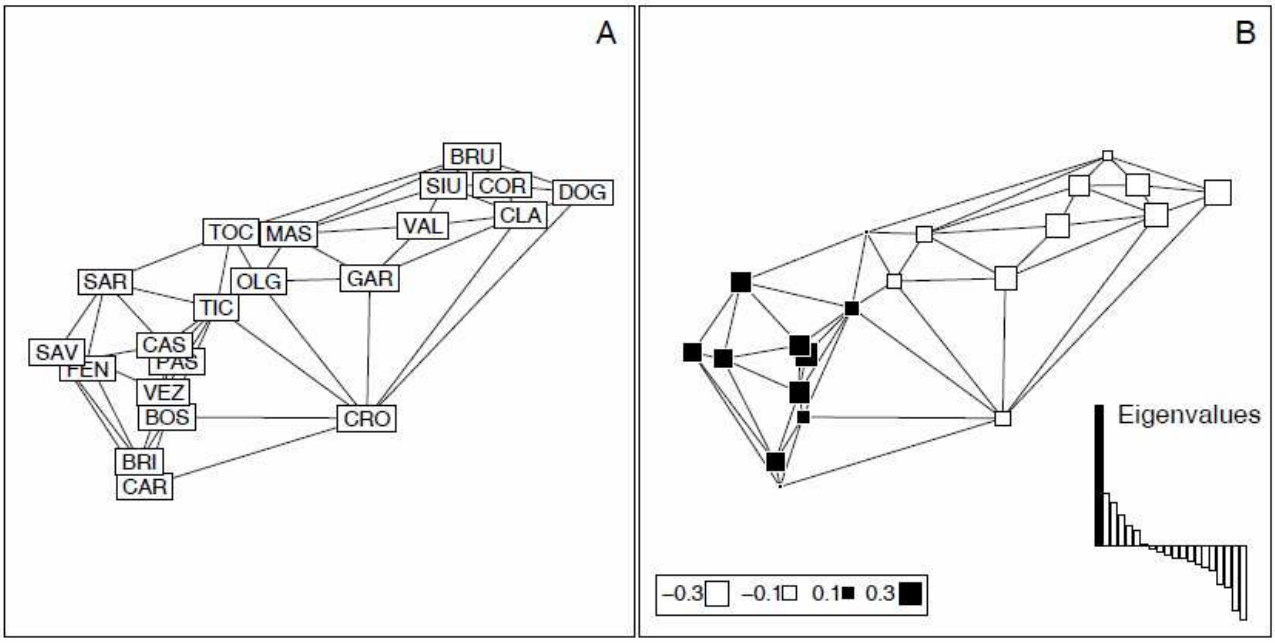
685

686 Figure 2



687

688 Figure 3



689