This is the peer reviewed version of the following article: Mayol, Maria, et al. "Adapting through glacial cycles : insights from a long-lived tree (Taxus baccata)" in New Phytologist, 2015, which has been published in final form at DOI: 10.1111/noh.13496. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

Adapting through glacial cycles: insights from a long-lived tree (*Taxus baccata* L.)

Maria Mayol*[1], Miquel Riba[1,2], Santiago C. González-Martínez[3], Francesca Bagnoli[4], Jacques-Louis de Beaulieu[5], Elisa Berganzo[1], Concetta Burgarella[6], Marta Dubreuil[1], Diana Krajmerová[7], Ladislav Paule[7], Ivana Romšáková[7], Cristina Vettori[8], Lucie Vincenot[1], Giovanni G. Vendramin[8]

1) CREAF, Cerdanyola del Vallès 08193, Spain
2) Univ Autonòma Barcelona, Cerdanyola del Vallès 08193, Spain
3) INIA, Forest Research Centre, Madrid 28040, Spain
4) Plant Protection Institute, National Research Council, Via Madonna del Piano 10, 50019 Sesto Fiorentino (FI), Italy
5) CNRS-UMR7263-IMBE, Université Paul Cézanne, Aix-en-Provence, France
6) Institut de Sciences de l'Evolution de Montpellier, Université Montpellier 2, CNRS UMR 5554
7) Faculty of Forestry, Technical University, SK-96053 Zvolen, Slovakia
8) Institute of Biosciences and Bioresources, National Research Council, Via Madonna del Piano 10, 50019 Sesto Fiorentino (FI), Italy

*Corresponding author: Phone: +34935814679; E-mail: Maria.Mayol@uab.es

Total word count: 6474
Summary: 194
Introduction: 905
Materials and Methods: 2018
Results: 1656
Discussion: 1548
Acknowledgements: 153
Figures: 5 (all figures should be published in color)
Tables: 3
Supporting information: 8 Figures; 6 Tables; 2 Notes.

Summary

- Despite the large body of research devoted to understand the role of Quaternary glacial cycles in genetic divergence of European trees, the differential contribution of geographic isolation and/or environmental adaptation in creating population genetic divergence remains unexplored. In this study, we used a long-lived tree (*Taxus baccata* L.) as a model species to investigate the impact of Quaternary climatic changes on genetic diversity via neutral (IBD) and selective (IBA) processes.
- We applied Approximate Bayesian Computation to genetic data to infer its demographic history, and combined this information with past and present climatic data to assess the role of environment and geography in the observed patterns of genetic structure.
- We found evidence that yew colonized Europe from the East, and that European samples diverged into two groups (*Western*, *Eastern*) at the beginning of the Quaternary glaciations, ~2.2 Myr before present (BP). Apart from the expected effects of geographical isolation during glacials, we discovered a significant role of environmental adaptation during interglacials at the origin of genetic divergence between both groups.
- This process may be common in other organisms, providing new research lines to explore the effect of Quaternary climatic factors on present-day patterns of genetic diversity.

Key words: Approximate Bayesian Computation; cpDNA; demography; evolutionary history; environment-dependent selection; interglacial; microsatellites; *Taxus baccata*.

Introduction

It is currently accepted that Quaternary climatic oscillations have played a major role in shaping the geographical distribution of European species and their patterns of genetic structure (Hewitt, 2004). In the particular case of temperate taxa, geographical isolation and long-term persistence in southern refugia during the glacial episodes have been considered essential for population divergence and the emergence of new lineages (Hampe & Petit, 2005). However, climatic conditions experienced during glacial and interglacial intervals could also have provided opportunities for genetic divergence through selective pressures and adaptation associated with different local or regional environments. In populations adapted to ecologically dissimilar habitats, gene flow can be limited by selection against maladapted immigrants (Nosil *et al*., 2005), and this might in turn have indirect effects on the whole genome, since reduction of gene flow promotes neutral divergence through increased genetic drift (Wright, 1931). In this case, genetic differentiation inferred from neutral markers is expected to be correlated with differences in local environments, a pattern that has been described as "Isolation-By-Adaptation" (IBA, Nosil *et al*., 2008), in analogy to standard patterns of genetic differentiation with geographical distance, i.e. "Isolation-By-Distance" (IBD).

Despite the large body of research devoted to understand the effects of climatic changes of the Quaternary, few studies to date have investigated the differential contribution of geographic isolation and/or climatic adaptation in creating population genetic divergence in temperate species. Long-lived organisms, such as trees, are especially well suited models to address these questions. Many temperate trees are distributed over large areas characterised by a wide heterogeneity of both biotic and abiotic factors, and show local adaptation to environmental gradients at multiple spatial scales (Savolainen *et al*., 2007), which can generate IBA patterns. Because of the buffering effects of their life history traits (great longevity, overlapping generations, prolonged juvenile phase) on changes in genetic structure (Austerlitz *et al*., 2000), long-lived trees offer additional advantages over short-lived organisms to investigate the generation of IBA along the Quaternary, allowing to explore how much genetic variation is associated with current or past environmental conditions. For instance, the effects of climate during the last glaciation are still evident on contemporary patterns of genetic variation of long-lived *Quercus engelmannii* (Ortego *et al*., 2012) and *Q. lobata* (Gugger *et al*., 2023), suggesting that the genetic signal of past climate can persist over

extended time periods in organisms with large effective population sizes and long generation times. Nevertheless, unravelling the effect of different climatic periods on spatial genetic divergence is challenging, because current observed patterns may result from the interplay among processes acting at different temporal scales. Assessing the most likely time course for the appearance of environmental barriers to gene flow is the first step to accurately dissect their role as actual contributors to IBA. Today, the existence of various palaeoclimatic databases allows to evaluate the effect of period-specific climatic conditions on neutral genetic diversity by testing each period separately, and recently developed Approximate Bayesian Computation (ABC) methods can be used to elucidate complex demographic scenarios with relatively low demands in terms of computation effort (Beaumont, 2010), as well as to estimate the time of the inferred demographic processes.

English yew (*Taxus baccata* L.) is a long-lived, slow growing Tertiary relict (Hao *et al*., 2008) native of Eurasian temperate and Mediterranean forests. Extending from North Africa to Scandinavia, and from the Iberian Peninsula to the Caspian Sea, yew grows under a wide range of environmental conditions, from oceanic to continental and Mediterranean climate (Thomas & Polwart, 2003). Although palaeoecological information on past yew distribution is scarce, some of the longest European Pleistocene pollen records indicate that *Taxus* expanded its range during several interglacials and made a much more significant contribution to vegetation in Europe than today (Mamakova, 1989; Turner, 2000; de Beaulieu *et al*., 2001; Müller *et al*., 2003; Koutsodendris *et al*., 2010). Palynological records also indicate that yew was able to persist during the last glaciation, not only in southern refugia (Allen *et al*., 2002; Carrión, 2002; Carrión *et al*., 2003), but also in Central and Eastern Europe (Stewart & Lister, 2001; Willis & van Andel, 2004), although some debate still exists on the presence of cryptic refugia in northern Europe (Tzedakis *et al*., 2013). The wide extent of environmental heterogeneity within the species' range, together with its long presence in Europe, make English yew an ideal species to investigate the impact of Quaternary climatic changes on genetic diversity via neutral and selective processes.

In this study, we used an integrated approach combining genetic and palaeoenviromental data to (1) elucidate the demographic history of *T. baccata* throughout its range, and (2) determined the role of environmental and geographical factors in generating the observed patterns of genetic structure. About five thousand trees from 238 localities covering yew natural range were genotyped with neutral

microsatellite markers to identify distinct genetic clusters. ABC was used to select the most likely scenario shaping genetic diversity in this species and to set an approximate time frame for the inferred history. Finally, we used the available climatic information for three time periods, i.e. the last interglacial (LIG, ~120,000-140,000 yrs BP), the last glacial maximum (LGM, ~21,000 yrs BP), and present conditions (PRE, ~1950-2000), to evaluate the relative importance of current and past climatic conditions on the observed patterns of genetic variation. Coupling these approaches helped to determine whether standing patterns of genetic divergence are the result of historical isolation or, alternatively, of local adaptation to ecologically differentiated areas.

Materials and Methods

Sampling, DNA extraction and nuclear microsatellite genotyping

A total of 4,992 samples (*N*=1-60 per locality, mean 21) were collected at 238 localities covering the entire distribution range of *Taxus baccata* L. (Fig. 1; Table S1). Total DNA was isolated from 50-100 mg of dry leaf material using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) or a modified protocol from Dellaporta *et al*. (1983). Seven primer pairs for the amplification of nuclear microsatellites (nuSSRs) developed specifically for *T. baccata* were used for the genetic analysis following conditions described in Dubreuil *et al*. (2008).

Chloroplast DNA sequencing

Six chloroplast regions were tested following the PCR conditions given in Shaw *et al*. (2005): *rbcL*, *rpl36-rps8*, *trnH-psbA*, *trnC-ycf6*, *trnT-trnL* and *trnL-trnF*. Additionally, the *trnS-trnQ* spacer region was amplified as in Schirone *et al*. (2010). Only three of them were successfully amplified: *rbcL*, *trnS-trnQ* and *trnL-trnF*. The amplified products were screened for polymorphism using 1-2 individuals from 18-26 populations (Tables S2, S3) sampled across the distribution range of the species. PCR products were purified using the QIAquick gel extraction kit (Qiagen, Hilden, Germany), and sequenced from both ends with an ABI 377 automated sequencer using the ABI BigDyeTM Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster City, California, USA). Sequences available from previous studies (Shah *et al*., 2008; Schirone *et al*., 2010) were downloaded from GenBank and aligned with the newly obtained sequences (accession numbers KP115899-KP115935).

Genetic diversity and structure based on nuclear microsatellites

In all populations with at least 8 individuals (195 locations, N=4,829), observed (H_O) and expected (H_E) heterozygosity were computed using GENETIX v.4.04 (Belkhir *et al*., 2001). The number of private alleles and allelic richness (A_R) were calculated using GENALEX v.6.5 (Peakall & Smouse, 2012) and FSTAT v.2.9.3.2 (Goudet, 2001), respectively. Linkage disequilibrium among all pairs of loci within each population was assessed by the Markov-chain approximation of the Fisher's exact test implemented in GENEPOP v.4.0.7 (Rousset, 2008).

The same 195 populations were used to investigate genetic structure using different approaches. AMOVA (Excoffier *et al*., 1992) was used to partition total molecular variance within and among populations using ARLEQUIN v.3.5.1.2 (Excoffier *et al*., 2005). Significance was obtained by non-parametric permutation using 10,000 replicates. Multi-locus F_{ST} was estimated correcting for the possible presence of null alleles with the program FREENA (Chapuis & Estoup, 2007), using 1,000 bootstraps to compute 95% confidence intervals. IBD was evaluated by testing the correlation between the matrix of pairwise $[F_{ST}/(1-F_{ST})]$ and the matrix of geographic distances (logarithmic scale) using the Mantel test implemented in GENETIX v.4.04, with 10,000 permutations. Finally, Jost's estimator D_{est} (Jost, 2008) was computed to assess population differentiation using GENALEX v.6.5, and significance was evaluated using 1,000 replicates.

Two Bayesian clustering methods were used to infer distinct gene pools within the full dataset (238 locations, N=4,992). The program STRUCTURE v.2.2 (Pritchard *et al*., 2000) was run without prior population information, selecting the correlated allele frequencies model and assuming admixture. Ten independent runs for each K cluster (from K=1 to 7) were performed, setting burn-in and run lengths of 50,000 and 500,000 iterations, respectively. The best number of clusters was determined following the recommendations by Pritchard & Wen (2004) and Evanno *et al*. (2005). Mean F_{ST} values measuring the divergence of each inferred cluster from a single hypothetical "ancestral population" were also obtained (Falush *et al*., 2003). Additionally, TESS v.2.3 (François *et al*., 2006) was employed to estimate the number of genetic clusters (K) present in the data incorporating the geographical location of individuals as a prior information. Both admixture (BMY model; spatial interaction parameter: 0.6) and non-admixture models were used to perform five independent runs for K ranging from 1 to

7, with a burn-in of 10,000 and a total number of 50,000 sweeps. The optimal value of *K* was determined by plotting the deviance information criterion (DIC) against *K* and choosing the values of K_{max} corresponding to a plateau of the curve (François & Durand, 2010). To graphically represent the results obtained, the averaged results of the assignments of STRUCTURE and TESS were plotted on maps generated with ARCGIS v.9.1. (ESRI, Redlands, California).

Demographic history (ABC models)

We applied the ABC framework implemented in DIYABC v.1.0.4.46 (Cornuet *et al*., 2010) to nuSSRs to infer the demographic history of *T. baccata*. The information obtained from the above analyses was the starting point for designing the different scenarios to test (for further details, see Results). STRUCTURE and TESS identified slightly different best number of clusters (*K*=2 (*Western*, *Eastern*) and *K*=3 (*Western*, *Eastern*, *Iran*), respectively), but both supported (i) a first partition between *Western* and *Eastern* samples, and (ii) the presence of *Admixed* populations at the intersection of both clusters, suggesting a secondary contact of divergent lineages. However, a clear westward cline of decreasing diversity from Iran to the Mediterranean area was also detected, which might be indicative of a colonization pattern. Thus, we designed four scenarios to test alternative hypotheses considering two or three genetic pools (Fig. 2). Scenario A tested a "secondary contact" of two separated gene pools (*Western*, *Eastern*). Scenarios B and D tested a "colonization" event from the east, the former considering two genetic pools (*Western*, *Eastern*), and the latter with three (*Western*, *Eastern*, *Iran*). Scenario C considered a "colonization" from *Iran*, the separation of European samples into two genetic pools (*Eastern*, *Western*), and a posterior "secondary contact".

For Scenarios A and B, three groups of populations were created: *Eastern*, *Western* and *Admixed*. A population was considered as admixed when the proportion of individuals assigned to the eastern or the western cluster was less than 70%. The proportion of membership and the assignment of populations to each specific group are reported in Table S1. Given that recent studies suggest that pooling data across populations can be a problem to infer demography (Chikhi *et al*., 2010), the following procedure was designed to avoid the potential confounding effects of population structure on the inference of demographic parameters. Ten different sets of populations were constructed, each containing approximately 500 individuals, i.e. representing

about 10% of the whole dataset (Fig. S1). Each set was composed by ~200 individuals belonging to the *Eastern* pool, ~200 individuals from the *Western* one, and ~100 of *Admixed* composition (hereafter called "500-sample datasets"). To minimize the effect of spatial genetic structure, we sought that the populations included in each dataset were geographically close or that genetic divergence among populations was low. For Scenarios C and D, two additional datasets with four groups of populations were constructed (Fig. S1), i.e. considering the *Iran* pool (Iran, Georgia) to be independent from the *Eastern* one, as inferred using TESS for K=3 and STRUCTURE for K=4. In this case, datasets were composed by ~200 individuals belonging to the *Eastern*, *Western* and *Iran* pools, respectively, and ~100 of *Admixed* composition (hereafter called "700-sample datasets"). The composition of these two datasets for each ABC run was different for all pools except for the *Iran* one, where only 199 individuals were available.

One million simulations were run for each dataset. Prior parameter distributions were chosen as broad as possible to explore a wide range of population sizes and time frames (measured in generations): Uniform [10; 100,000] for current effective population sizes, Uniform [1; 100,000] for divergence times t1, t2 and t3 (with t3>t2 and t2>t1), Uniform [10; 1,000,000] for ancestral effective population size, and Uniform [0.001; 0.999] for admixture rate. A generalized stepwise mutation model was assumed and default values were used for all prior mutation parameters, except for the mean mutation rate, for which minimum and maximum default values (10^{-4}-10^{-3} mutations per locus per generation) were enlarged to 10^{-5}-10^{-3} after previous runs giving biased posteriors towards the lower mutation rate. Each simulation was summarized by the following statistics: mean number of alleles and mean genetic diversity (Nei, 1987) for each cluster, and mean number of alleles, mean genetic diversity, F_{ST}, mean index of classification (Rannala & Mountain, 1997), and shared allele distance (Chakraborty & Jin, 1993) between pairs of clusters. After ensuring that this combination of scenarios and priors was able to produce datasets similar to the observed one (Fig. S2), the posterior probabilities of each scenario were calculated with a local logistic regression procedure using the 1% closest simulated points. Retained simulations were used to infer parameter posterior distributions by local linear regression using a logit transformation of the parameters. The reliability of the model and chosen scenario was evaluated for each of the twelve simulations by performing model checking and computing the confidence in scenario choice (see Notes S1 for further details).

Past and present impact of environmental factors on genetic structure

We used the climatic information available at the WorldClim database (Hijmans *et al*., 2005) to evaluate the effect of past and present climatic conditions on current genetic structure. For the present time (PRE, ~1950-2000), nineteen bioclimatic variables were downloaded for the 195 populations with $N \geq 8$. Two bioclimatic variables that were highly correlated with the others ($r > 0.9$) were excluded, and the remaining variables were summarized into the first two axes of a Principal Component Analysis (PCA) using R (R Core Team, 2013). The environmental variables with loadings on the PCA axes higher than 0.5 (Table S4) and the 238 occurrence points were used to model current climatically suitable areas for English yew with maximum entropy (MAXENTv.3.3.3) and BIOCLIM (DIVA-GIS v.7.5.) algorithms. Predictions were also generated separately for the *Western* (153 sampling sites) and *Eastern* (64 sampling sites) gene pools to examine whether genetic divergence among them was environmentally induced. The modelled distributions were generated with 75% of the points (training data) and cross-validated with 25% of the remaining localities (test data), averaged over ten runs. The performance of the models was tested by measuring the area under the Receiver Operating Characteristic curve (AUC). The logistic outputs of MAXENT models were transformed to presence-absence maps using the maximum training sensitivity plus specificity (MTSS) threshold. For BIOCLIM, maps were obtained leaving only values with high, very high or excellent suitability (i.e., within the 5-95th percentile interval).

To determine the contribution of present environment on genetic differentiation, we tested for the relationship between pairwise F_{ST} and climatic distance while controlling for geographic distance. We computed climatic (Euclidian) distance matrices based on population scores for both PCA axes (PC1, PC2), and for each environmental variable. Tests were performed for the whole dataset and for each genetic cluster (*Eastern*, *Western*) using partial Mantel tests ("mantel.partial" function; R Core Team, 2013) and Multiple Matrix Regressions (MMRR script; Wang, 2013). Significance tests were based on 10,000 permutations. To reduce the risk of spurious correlations, in particular for less conservative MMRR tests, we only considered those correlations that were significant with both methods.

The same procedures were applied to investigate the contribution of past climate to current genetic differentiation. We projected the models for the present onto three

paleoclimate layers, the Community Climate System Model (CCSM) and the Model for Interdisciplinary Research on Climate (MIROC) for the last glacial maximum (LGM, ~21,000 yrs BP), and the model for the last interglacial (LIG, ~120,000-140,000 yrs BP). Then, to investigate the correlation of past climate and observed genetic differentiation (F_{ST}), we retained only those populations where suitable environment have existed for yew persistence during these periods. Although distribution of *T. baccata* may have not been exactly the same across the Quaternary, projections suggest a rather stable distribution of the species in relatively large parts of its range (see Results), so using only populations that could have been located at or near the present locations can be considered a reasonable approximation of the distribution of the species in the past. Thus, we selected those occurrence points with logistic output values in MAXENT above the respective MTSS thresholds in each model, and with suitability values above the 5-95th percentile interval in BIOCLIM. Because of the present distribution of English yew was more accurately predicted when combining modelled distributions of single gene pools than from the full model (see Results), we constructed datasets combining the predicted suitable populations obtained for *Western* and *Eastern* models of past climate (indicated in Table S1), and used them to perform partial Mantel and MMRR tests as for the present time.

Results

Chloroplast DNA sequencing

Chloroplast regions comprised 1,359, 656 and 344 aligned positions for *rbcL*, *trnS-trnQ* and *trnL-trnF*, respectively. No polymorphism was found for the *rbcL* gene, and *trnS-trnQ* and *trnL-trnF* showed two closely related haplotypes, respectively (Tables S2, S3). For both markers, populations harbouring different haplotypes were located at Guilan and Golestan provinces (Iran), at the eastern extreme of the distribution of English yew (Fig. S3).

Genetic diversity and structure based on nuclear microsatellites

Allelic richness (A_R), expected (H_E) and observed (H_O) heterozygosity ranged from 2.243 to 5.295, 0.354 to 0.855, and 0.171 to 0.768, respectively (Table S1). Among a total of 3,957 tests for linkage disequilibrium between pairs of loci, 98 were significant (P<0.05) after sequential Bonferroni corrections, but almost all involved the same eight

populations. Since the application of Bayesian methods rely on the assumption of linkage equilibrium between loci, we performed additional runs with STRUCTURE program excluding these populations.

Very similar pairwise F_{ST} were obtained when correcting or not for the presence of null alleles. Corrected values ranged from 0.001 to 0.599, with an overall F_{ST}=0.149. Only 84 out of 18,915 population pairs were not significantly (P<0.05) differentiated from each other after a sequential Bonferroni correction for multiple tests (Table S5). These results were in agreement with AMOVA, with a 16.41 % of the total variance explained by differences among populations (Table S6). Overall Jost's D_{est} was 0.478, indicating that the proportion of allelic differentiation among populations was higher than the proportion of variance in allele frequencies. The correlation between genetic and geographic distances was highly significant (r=0.281, P<0.001), suggesting the existence of an isolation-by-distance pattern.

STRUCTURE runs including or excluding populations with significant linkage between loci produced almost identical results. The method identified an optimal partition in two genetic pools with a clear geographical pattern: populations from central Europe to Iran (*Eastern*) and populations from the western range (*Western*), with a contact zone of admixed populations (*Admixed*) located along Central Europe, Italy and the Mediterranean islands (Fig. 3). An additional partition (K=3) subdivided the western group into two differentiated pools, predominantly located in Central Europe and the Mediterranean area, respectively (Fig. S4). Increasing the number of partitions (K=4) produced the splitting of samples from Iran and Georgia as an independent pool (*Iran*) within the *Eastern* one (Fig. S4). The F_{ST} values obtained by the correlated frequencies model in STRUCTURE increased toward the west, suggesting that eastern populations were closer to the hypothetical "ancestral population".

The best model with TESS was the one considering admixture, and generated a very similar population clustering for K=2 (Fig. S4), but the best number of clusters was inferred at K=3 (Fig. 3), and the splitting of the easternmost group (*Iran*) occurred earlier (i.e. for K=3 in TESS, and for K=4 in STRUCTURE). Nevertheless, the same trends were identified with both approaches: (i) the first level of divergence produced the partition of *Western* and *Eastern* samples, (ii) populations showing higher levels of admixture were located at the intersection of both clusters, (iii) divergence from the hypothetical "ancestral population" increased towards the west, and (iv) increasing

number of partitions led to the split of the easternmost samples (Georgia, Iran) as an independent group (*Iran*).

Mean genetic diversity of populations assigned to the *Eastern* pool and those of *Admixed* composition was significantly higher than that of the *Western* cluster (mean $A_{R(E)}$=4.46, $A_{R(A)}$=4.36, $A_{R(W)}$=3.61; mean $H_{E(E)}$=0.773, $H_{E(A)}$=0.759; $H_{E(W)}$=0.653; Duncan's test after ANOVA: P<0.001), indicating a pattern of decreasing genetic diversity from east to west (Fig. 4). In addition, the number of populations displaying private alleles was higher in the *Eastern* pool than in the *Western* one (23 vs. 15), as well as the number of private alleles (43 vs. 17). Of these 43 private alleles, 23 were detected in populations from Iran and Georgia (*Iran* pool). Only 7 *Admixed* populations had private alleles, and the proportion was low (7 out of 67). Jost's estimator indicated higher differences in allele composition among populations in the east ($D_{est(E)}$=0.489, $D_{est(A)}$=0.387, $D_{est(W)}$=0.351), while greater deviations from panmixia (F_{ST}) were detected in the west ($F_{ST(E)}$=0.127, $F_{ST(A)}$=0.103, $F_{ST(W)}$=0.147).

Demographic history (ABC models)

All the ABC simulations were able to discriminate between the tested scenarios, with high posterior probabilities and 95% confidence intervals never overlapping those of the other scenario (Table 1). The most likely scenario using the "500-sample datasets" was Scenario A, with a strong support in almost all cases (PP≥0.8; Table 1). However, scenario B was always chosen when the datasets used for simulations included the easternmost samples (Georgia, Iran) as representatives of the *Eastern* pool (sim4, sim5 and sim8, see Fig. S1), albeit with lower support (Table 1). This was in accordance with genetic diversity distribution and, together, suggest an eastern colonization of Europe, with easternmost populations constituting an independent gene pool (*Iran*) that was not the source of admixture in Central Europe. Similarly, simulations performed on the two "700-sample datasets", i.e. considering three differentiated pools (*Iran*, *Eastern*, *Western*), unambiguously indicated support for Scenario C (>0.9; Table 1), which tested a first migration wave from the east (*Iran*), a more recent separation of European samples in two distinct pools (*Eastern*, *Western*), and a secondary contact (*Admixed*) between them (Fig. 2). Model testing procedures further supported the reliability of this scenario (see Notes S1 for further information). Parameter posterior distributions are shown in Fig. S5.

Under this model (Scenario C), estimated time of divergence among Iranian and European samples would have occurred, on average, about 6 Myr BP (90% credible intervals: 1.35-14.78 Myr BP, Table 1), assuming a generation time of ~100 years for English yew. Although reproduction can begin earlier when growing under open canopy conditions, yew usually grows as isolated understory tree, and reach maturity later, between 70-120 yr (Thomas & Polwart, 2003; L. Paule & M. Riba, unpublished data). The posterior separation of *Eastern* and *Western* clusters, and the subsequent admixture event would have taken place about 2.2 Myr BP (90% credible intervals: 0.5-7.5 Myr BP) and 200,000 years BP (90% credible intervals: 50,000-800,000 years BP), respectively (Table 1). These estimates are mostly in agreement with those obtained with the "500-sample datasets", with averaged estimates of both events around 2 Myr BP and 230,000 years BP, respectively (Table 1). Although 90% credible intervals are large for all the simulations, model checking confirmed that the model was consistent with the observed data, suggesting that large confidence intervals are due to data information content and not to a model misfit (Notes S1).

Past and present impact of environmental factors on genetic structure

The first two PCA axes explained 52% of the variation for the present climate. PC1 was mainly correlated with temperatures, whereas PC2 was positively correlated with precipitation (Table S4). Despite some overlap, populations belonging to *Western* and *Eastern* clusters defined clear groups along the first axis of the multivariate space (Fig. S6), indicating that substantial differences exist in climate for each geographic region. On average, populations within the *Western* cluster experienced smaller seasonal temperature fluctuations, warmer temperatures (annual mean, minimum and maximum), higher temperatures during the driest quarter, and less precipitation during the warmest quarter (ANOVA: $P<0.001$).

Averaged AUC values for the replicate runs were >0.870 for all distribution models, supporting their predictive power. The predicted full species model for the present generated with MAXENT was fairly congruent with yew current distribution, (Fig. S7), but this was more accurately predicted when combining modelled distributions of single gene pools, especially with regard to Eastern Europe (Fig. 5). Similar results were obtained using the BIOCLIM algorithm (Notes S2).

The CCSM and MIROC models for the LGM yielded large differences in predicted distributions, and were highly dependent on the algorithm used (Figs. 5, S7;

Notes S2). MAXENT models suggested much wider suitable areas than BIOCLIM, especially with regard to CCSM. However, all models supported the existence of large suitable areas for English yew in several southern refugia (i.e., the Balkans, Iberia and Italy). Projections for LIG produced similar models with MAXENT and BIOCLIM, showing a westward shift with respect to present-day climatic conditions, both for the *Western* and *Eastern* clusters (Figs. 5, S7; Notes S2).

Despite a substantial lack of precision for LGM models, a common trend was that models produced using localities from either *Western* or *Eastern* gene pools alone showed little overlap of their predicted distributions for all periods considered, especially during both interglacials (Fig. 5), suggesting that *Eastern* and *Western* clusters have occupied environmentally different regions since the long past.

After controlling for geographic distance, and at the scale of the whole species range, there was a significant positive association between pairwise F_{ST} and the PC1 axis for the present climate ($r_{Env\text{-}PRE}$=0.157, $b_{Env\text{-}PRE}$=0.154, P<0.001), while relations with PC2 variables were not significant (Table 2). Analysed separately, annual mean temperature ($r_{Env\text{-}PRE}$=0.171, $b_{Env\text{-}PRE}$=0.161, P<0.001) and minimum temperature of the coldest month ($r_{Env\text{-}PRE}$=0.141, $b_{Env\text{-}PRE}$=0.145, P<0.001) were the most relevant variables explaining population genetic structure. Very similar results were found within the *Western* pool, with a significant correlation among genetic differentiation and PC1 variables ($r_{Env\text{-}PRE}$=0.174, $b_{Env\text{-}PRE}$=0.175, P<0.01), and more specifically with annual mean temperature ($r_{Env\text{-}PRE}$=0.219, $b_{Env\text{-}PRE}$=0.218, P<0.01) and minimum temperature of the coldest month ($r_{Env\text{-}PRE}$=0.163, $b_{Env\text{-}PRE}$=0.165, P<0.01). Within the *Eastern* pool, no significant correlations were found for both tests except for temperature seasonality ($r_{Env\text{-}PRE}$=0.136, $b_{Env\text{-}PRE}$=0.150, P<0.05).

MAXENT projections onto past climatic models suggested that suitable conditions would have existed for the persistence of 102-123 and 94 populations during LGM and LIG, respectively (Table 3). In these populations, we found that LIG climate contributed similarly as the present to genetic divergence, with a positive association between F_{ST} and annual mean temperature ($r_{Env\text{-}LIG}$=0.104, $b_{Env\text{-}LIG}$=0.115, P<0.05), and minimum temperature of the coldest month ($r_{Env\text{-}LIG}$=0.096, $b_{Env\text{-}LIG}$=0.106, P<0.05). In addition, we found a significant contribution of mean diurnal temperature range ($r_{Env\text{-}LIG}$=0.177, $b_{Env\text{-}LIG}$=0.170, P<0.01), and no significant correlation with the PC2 axis (Table 3). Positive significant associations for both Mantel and MMRR tests were not

detected during LGM models. These results were confirmed when using suitable populations predicted with BIOCLIM (Notes S2).

Discussion

Demographic history of English yew

The combination of Bayesian clustering and Approximate Bayesian Computation methods shed light on the demographic history of English yew. According to our ABC results, nuSSRs in *Taxus* seem to retain the imprint of very ancient events, as suggested by the divergence time estimates for the inferred demographic processes. Even so, a near absence of variation and spatial structure for chloroplast DNA markers was observed, in accordance with the slow chloroplast nucleotide substitution rate reported in conifers (Willyard *et al*., 2007).

ABC simulations suggest that the most likely demographic scenario for *T. baccata* involves a first migration wave from eastern territories (*Iran*) to the west, a more recent separation of the European samples into two gene pools (*Eastern*, *Western*), and a secondary contact (*Admixed*) of both clusters along Central Europe, Italy and Mediterranean islands (Fig. 2, Scenario C). The ancient migration from the east is also supported by the westward decline of genetic diversity (Fig. 4), and by the fact that F_{ST} values from the hypothetical "ancestral population" for each inferred cluster always increased towards the west, suggesting that easternmost populations were closer to the ancestral one. In agreement with our results, recent studies set the origin of *Taxus* in North America or South West China during the late Cretaceous to mid Eocene (66.55 ± 11.22 Myr BP), from which was dispersed to the current distribution areas (Hao *et al*., 2008). The genus probably reached Europe through the Irano-Turanian region, which has been postulated as a key source for the colonization of the Mediterranean region (Thompson, 2005; Mansion *et al*., 2008). This event probably occurred before the Lower Miocene, as indicated by the oldest fossil record (16-23 Myr BP; Kunzmann & Mai, 2005). An eastern origin and westward colonization of the Mediterranean, still reflected in current genetic structure, has also been postulated for other tree genera (e.g. *Abies*, Linares, 2011; *Frangula*, Petit *et al*., 2005; *Laurus*, Rodríguez-Sánchez *et al*., 2009).

Our ABC simulations place the separation among the Iranian and European genetic pools around 6 Myr BP, although 90% credible intervals are wide (on average

[1.35-14.78] Myr BP, Table 1). In addition, DIYABC does not model continuous gene flow at each generation, which could have led to underestimating the divergence time. Nevertheless, assuming lack or reduced gene flow is a reasonable assumption in English yew (see Dubreuil *et al*., 2010; González-Martínez *et al*., 2010; Chybicki *et al*., 2011; Burgarella *et al*., 2012), as also suggested by high levels of pairwise genetic differentiation in our study (see Results). An ancient separation is also evident from the results obtained using chloroplast DNA markers, since the only distinct haplotypes were located at the eastern extreme of the distribution (Guilan and Golestan provinces, Iran), suggesting that both groups became isolated long time ago. This is additionally confirmed by the significantly higher number of private alleles detected at nuclear microsatellites within the *Eastern* pool, particularly in populations from Iran and Georgia. This ancient vicariance might be associated to the intense changes that occurred during the Latest Miocene (6.1-5.7 Myr BP; Popov *et al*., 2006), which could have favoured both migration and differentiation within the Mediterranean Basin.

Around two million years before the present (90% credible intervals ~[0.5-7.5] Myr BP), the European populations split into two distinct genetic pools (*Eastern*, *Western*), and a posterior admixture of both lineages seem to have occurred about 200,000 years ago (90% credible intervals ~[50,000-800,000] years). These results are consistent with the expected pattern assuming that *T. baccata* survived in two allopatric refugia since the beginning of the Quaternary, from which they expanded and converged further north during warm interglacial periods. Such an east-west pattern of differentiation across the Mediterranean region has been reported for other trees (*Laurus nobilis,* Rodríguez-Sánchez *et al*., 2009; *Olea europaea*, Besnard *et al*., 2007; *Quercus suber*, Lumaret *et al*., 2002), herbaceous species (*Arabidopsis thaliana,* François *et al*., 2008) and coastal plants (*Carex extensa*, Escudero *et al*., 2010), and has been interpreted as a result of an east-west isolation during glaciations of the Quaternary (e.g., François *et al*. 2008, Escudero *et al*. 2010). Our results, however, suggest that interglacials could have played a key role in maintaining genetic divergence between both groups, as we discuss in the next section.

Past and present impact of environmental factors on genetic structure

Our analyses support the hypothesis that both geography and climate have played a significant role in shaping genetic structure of English yew. The divergence between *Western* and *Eastern* clusters can be explained by their persistence in spatially isolated

refugia during glacial periods. However, species distribution models (Fig. 5) revealed an almost non-overlapping distribution of both groups linked to distinct climatic regimes, particularly during interglacials, which may have reinforced the divergence of the two lineages through differential adaptation to their respective environments. This was in accordance to the results of partial Mantel tests and MMRRs, showing significant positive correlations for the present interglacial between genetic distance and temperature variables, both when considering the whole species range or the *Western* and *Eastern* samples separately (Table 2). Similar correlations were also found for the last interglacial, with temperature variables remaining as significant predictors of genetic distance after accounting for geography at the species level (Table 3). During the last glacial maximum, however, we did not find any significant positive association between climate and present-day patterns of genetic differentiation (Table 3).

Contrary to previous ecological studies highlighting the importance of water availability on *T. baccata* demographic processes, such as regeneration success (Sanz *et al*., 2009) or population sex ratio (Iszkuło *et al*., 2009), our results did not reveal a direct effect of rainfall variables (PC2) on genetic divergence, but rather pointed to a major effect of the temperature. The importance of temperature as a selective agent has been well documented in several tree species, usually linked to altitudinal, latitudinal or longitudinal clines (e.g. Jump *et al*., 2006; Grivet *el al*., 2011; Prunier *et al*., 2013). Even though the IBA patterns detected in this study do not imply causality and selection cannot be explicitly tested with our current data, the importance of temperature as a selective agent on English yew is supported by common garden observations, where significant regional differences associated with temperature clines are found in growth and phenology (own unpublished data). Moreover, in a study comparable in scale to the present work (92 populations), we found a significant association between sex-ratio and temperature, but western (Western Mediterranean and British Isles) and eastern (Central and Northern Europe) populations were clearly clustered into two distinct groups (see Fig. S8, after Berganzo, 2009), suggesting the existence of two evolutionary lineages adapted to contrasted temperature ranges. This gives additional support to the role of climate-driven adaptation in the divergence of *Eastern* and *Western* groups after initial isolation in allopatric refugia.

Several studies have reported the joint influence of isolation by distance and environmental adaptation to promote genetic divergence of plant populations (Lee & Mitchell-Olds, 2011; Temunović *et al*., 2012; Mosca *et al*., 2013). For example, both

geological and climatic changes during the Pliocene and Pleistocene has been proposed to explain the divergence of lineages of some conifers of the Qinghai-Tibet Plateau, such as *Taxus wallichiana* (Liu *et al*., 2013) of *Picea likiangensis* (Li *et al*., 2013). Nevertheless, none of them have reported evidence of a differential contribution of warm and cold periods of the Quaternary in generating genetic divergence of populations or groups through geographic isolation and/or climatic adaptation. Our results suggest that environmental factors during warm interglacials could have been crucial in shaping genetic variation of English yew. The correlation of LIG climate with present genetic variation also supports that the effects of past climate on genetic variation can persist for many generations, as already reported for other long-lived trees (Ortego *et al*., 2012; Gugger *et al*., 2023). However, unravelling the exact contribution of different interglacials (PRE, LIG) on environmentally-driven isolation is challenging, mainly because of the absence of an extensive fossil record. Although we cannot discard temporally varying selection, there is some evidence suggesting that adaptive processes would most likely have occurred during the last interglacial. Palaeoecological records indicate that English yew was much more abundant than today during interglacials preceding the last glaciation (e.g., Turner, 2000; de Beaulieu *et al*., 2001; Koutsodendris *et al*., 2010). After the Eemian (~115,000-130,000 yrs BP), *Taxus* is generally scarce in most of the European pollen records, suggesting a strong and continuous decline in its distribution. Molecular data also support strong reductions in effective population size starting between 100,000-300,000 yrs BP, and continuing up to the present in the Iberian Peninsula (Burgarella *et al*., 2012). Since large effective population sizes are expected to favor selection processes in relation to drift (Kimura *et al*., 1963; Charlesworth, 2009), environment-driven adaptation seem to be more likely in the past, when larger effective population sizes of *T. bacata* would have enhanced the efficiency of selection.

In conclusion, our results provide a distinct perspective for the climatic impact of Quaternary glaciations, suggesting that, despite being substantially shorter, selective pressures during interglacials could have had additional impacts on population genetic divergence to those of (extensively reported) geographical isolation during glacial periods. This opens new lines of research to explore the effect of Quaternary climatic factors on the present-day patterns of genetic diversity in other long-lived organisms.

Acknowledgements

We acknowledge L. Akzell, G. Bacchetta, D. Ballian, J. Bodziarczyk, "Bany-Al-Bahar Association", R. Brus, R. Crampton, L. Curtu, I.V. Delehan, X. Domene, A. El Boulli, A. Gailis, J. Gamisans, J. Gračan, P.C. Grant, D. Grivet, A. Harfouche, M. Heuertz, T. Hills, E. Imbert, G. Iszkuło, J. Kleinschmit, R. Klumpp, M. Konnert, E. Križová, T. Maaten, J. Mánek, M. Mardi, P. Mertens, T. Myking, M. Pakalne, M. Pridnya, I. Olivieri, B. Revuelta, J.A. Rosselló, G. Samuelsson, N. Shakarishvili, M. Sułkowska, E. Tessier du Cros, P.A. Thomas, U. Tröber, I. Tvauri, K. Ujházy, M. Valbuena-Carabaña, Ľ. Vaško, S. de Vries, N. Wahid, M. Zabal-Aguirre, V. Zatloukal and P. Zhelev for field assistance or providing yew samples. We also thank Michele Bozzano for draft shapefiles of yew distribution. This work was supported by grants CGL2007-63107/BOS, CGL2011-30182-C02-02, CSD2008-00040, 2009SGR608, VEGA1/3262/06, VEGA1/0745/09 and RBAP10A2T4. Part of the dataset presented here was included in I. Romšáková's PhD thesis.

References

Allen JRM, Watts WA, McGee E, Huntley B. 2002. Holocene environmental variability – the record from Lago Grande di Monticchio, Italy. *Quaternary International* **88:** 69-80.

Austerlitz F, Mariette S, Machon N, Gouyon P-H, Godelle B. 2000. Effects of colonization processes on genetic diversity: differences between annual plants and tree species. *Genetics* **154:** 1309-1321.

de Beaulieu JL, Andrieu-Ponel V, Reille M, Grüger E, Tzedakis C, Svobodova H. 2001. An attempt at correlation between the Velay pollen sequence and the Middle Pleistocene stratigraphy from central Europe. *Quaternary Science Reviews* **20:** 1593-1602.

Beaumont MA. 2010. Approximate Bayesian Computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics* **41:** 379-406.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F. 2001. GENETIX 4.04, logiciel sous Windows™ pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier, France.

Berganzo E. 2009. *Gender variation in Taxus baccata: the combined effects of stochasticity and sex-linked climate sensitivity*. MSc Thesis, Autonomous University of Barcelona, Barcelona, Spain.

Besnard G, Rubio de Casas R, Vargas P. 2007. Plastid and nuclear DNA polymorphism reveals historical processes of isolation and reticulation in the olive tree complex (*Olea europaea*). *Journal of Biogeography* **34:** 736-752.

Burgarella C, Navascués M, Zabal-Aguirre M, Berganzo E, Riba M, Mayol M, Vendramin GG, González-Martínez SC. 2012. Recent population decline and selection shape diversity of taxol-related genes. *Molecular Ecology* **21:** 3006-3021.

Carrión JS. 2002. Patterns and processes of Late Quaternary environmental change in a montane region of southwestern Europe. *Quaternary Science Reviews* **21:** 2047-2066.

Carrión JS, Yll EI, Walker MJ, Legaz AJ, Chaín C, López 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 & Biogeography* **12:** 119-129.

Chakraborty R, Jin L. 1993. A unified approach to study hypervariable polymorphisms: statistical considerations of determining relatedness and population distances. In: Pena SDJ, Chakraborty R, Epplen JT, Freys AJ, eds. *DNA fingerprinting: state of the science*. Birkhauser, Basel. 153-l 75.

Chapuis MP, Estoup A. 2007. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* **24:** 621-631.

Charlesworth B. 2009. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* **10:** 195-205.

Chikhi L, Sousa VC, Luisi P, Goossens B, Beaumont MA. 2010. The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics* **186:** 983-995.

Chybicki IJ, Oleksa A, Burczyk J. 2011. Increased inbreeding and strong kinship structure in *Taxus baccata* estimated from both AFLP and SSR data. *Heredity* **107:** 589-600.

Cornuet JM, Ravigné V, Estoup A. 2010. Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics* **11:** 401.

Dellaporta SL, Wood J, Hicks JB. 1983. A plant DNA minipreparation: Version II. *Plant Molecular Biology Reporter* **1:** 19-21.

Dubreuil M, Riba M, González-Martínez SC, Vendramin GG, Sebastiani F, Mayol M. 2010. Genetic effects of chronic habitat fragmentation revisited: strong genetic structure in a temperate tree, *Taxus baccata* (Taxaceae), with great dispersal capability. *American Journal of Botany* **97:** 303-310.

Dubreuil M, Sebastiani F, Mayol M, González-Martínez SC, Riba M, Vendramin GG. 2008. Isolation and characterization of polymorphic nuclear microsatellite loci in *Taxus baccata* L. *Conservation Genetics* **9:** 1665-1668.

Escudero M, Vargas P, Arens P, Ouborg NJ, Luceño M. 2010. The east-west-north colonization history of the Mediterranean and Europe by the coastal plant *Carex extensa* (Cyperaceae). *Molecular Ecology* **19:** 352-370.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14:** 2611-2620.

Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1:** 47-50.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131:** 479-491.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567-1587.

François O, Ancelet S, Guillot G. 2006. Bayesian clustering using hidden Markov random fields in spatial population genetics. *Genetics* **174:** 805-816.

François O, Blum MGB, Jakobsson M, Rosenberg NA. 2008. Demographic history of European populations of *Arabidopsis thaliana*. *PLoS Genetics* **4:** e1000075.

François O, Durand E. 2010. Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources* **10:** 773-784.

González-Martínez SC, Dubreuil M, Riba M, Vendramin GG, Sebastiani F, Mayol M. 2010. Spatial genetic structure of *Taxus baccata* L. in the western Mediterranean Basin: Past and present limits to gene movement over a broad geographic scale. *Molecular Phylogenetics and Evolution* **55:** 805-815.

Goudet J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from: http://www2.unil.ch/popgen/softwares/fstat.html). Updated from Goudet (1995).

Grivet D, Sebastiani F, Alía R, Bataillon Th, Torre S, Zabal-Aguirre M, Vendramin GG, González-Martínez SC. 2011. Molecular footprints of local adaptation in two Mediterranean conifers. *Molecular Biology and Evolution* **28:** 101-116.

Gugger PF, Ikegami M, Sork VL. 2013. Influence of late Quaternary climate change on present patterns of genetic variation in valley oak, *Quercus lobata* Née. *Molecular Ecology* **22:** 3598-3612.

Hampe A, Petit RJ. 2005. Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters* **8:** 461-467.

Hao DC, Xiao PG, Huang BL, Ge GB, Yang L. 2008. Interspecific relationships and origins of Taxaceae and Cephalotaxaceae revealed by partitioned Bayesian analyses

of chloroplast and nuclear DNA sequences. *Plant Systematics and Evolution* **276:** 89-104.

Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B-Biological Sciences* **359:** 183-195.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25:** 1965-1978.

Iszkuło G, Jasińska AK, Giertych MJ, Boratyński A. 2009. Do secondary sexual dimorphism and female intolerance to drought influence the sex ratio and extinction risk of *Taxus baccata*? *Plant Ecology* **200:** 229-240.

Jost L. 2008. G_{ST} and its relatives do not measure differentiation. *Molecular Ecology* **17:** 4015-4026.

Jump AS, Hunt JM, Martínez-Izquierdo JA, Peñuelas J. 2006. Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*. *Molecular Ecology* **15:** 3469-3480.

Kimura M, Maruyama T, Crow JF. 1963. The mutation load in small populations. *Genetics* **48:** 1303-1312.

Koutsodendris A, Müller UC, Pross J, Brauer A, Kotthoff U, Lotter AF. 2010. Vegetation dynamics and climate variability during the Holsteinian interglacial based on a pollen record from Dethlingen (northern Germany). *Quaternary Science Reviews* **29:** 3298-3307.

Kunzmann L, Mai DH. 2005. Die Koniferen der Mastixioideen-Flora von Wiesa bei Kamenz (Sachsen, Miozän) unter besonderer Berücksichtigung der Nadelblätter. *Palaeontographica* B **272:** 67-135.

Lee C-R, Mitchell-Olds T. 2011. Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Molecular Ecology* **20:** 4631-4642.

Li L, Abbott R, Liu B, Sun, Li L, Zou J, Wang X, Miehe G, Liu J. 2013. Pliocene intraspecific divergence and Plio-Pleistocene range expansions within *Picea likiangensis* (Lijiang spruce), a dominant forest tree of the Qinghai-Tibet Plateau. *Molecular Ecology* **22:** 5237-5255.

Linares JC. 2011. Biogeography and evolution of *Abies* (Pinaceae) in the Mediterranean Basin: the roles of long-term climatic change and glacial refugia. *Journal of Biogeography* **38:** 619-630.

Liu J, Möller M, Provan J, Gao L-M, Chandra Poudel R, Li D-Z. 2013. Geological and ecological factors drive cryptic speciation of yews in a biodiversity hotspot. *New Phytologist* **199:** 1093-1108.

Lumaret R, Mir C, Michaud H, Raynal V. 2002. Phylogeographical variation of chloroplast DNA in holm oak (*Quercus ilex* L.). *Molecular ecology* **11:** 2327-2336.

Mamakova K. 1989. Late Middle Polish glaciation, Eemian and Early Vistulian vegetation at Imbramowice near Wroclaw and the pollen stratigraphy of this part of the Pleistocene in Poland. *Acta Palaeobotanica* **29:** 11-176.

Mansion G, Rosenbaum G, Schoenenberger N, Bacchetta G, Rosselló JA, Conti E. 2008. Phylogenetic analysis informed by geological history supports multiple, sequential invasions of the Mediterranean Basin by the angiosperm family Araceae. *Systematic Biology* **57:** 269-285.

Mosca M, González-Martínez SC, Neale DB. 2014. Environmental versus geographical determinants of genetic structure in two subalpine conifers. *New Phytologist* **201:** 180-192.

Müller UC, Pross J, Bibus E. 2003. Vegetation response to rapid climate change in Central Europe during the past 140,000 yr based on evidence from the Füramoos pollen record. *Quaternary Research* **59:** 235-245.

Nei M. 1987. *Molecular Evolutionary Genetics*. New York, USA: Columbia University Press.

Nosil P, Egan SP, Funk DJ. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. *Evolution* **62:** 316-336.

Nosil P, Vines TH, Funk DJ. 2005. Perspective: reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* **59:** 705-719.

Ortego J, Riordan EC, Gugger PF, Sork VL. 2012. Influence of environmental heterogeneity on genetic diversity and structure in an endemic southern Californian oak. *Molecular Ecology* **21:** 3210-3223.

Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics* **28:** 2537-2539.

Petit RJ, Hampe A, Cheddadi R. 2005. Climate changes and tree phylogeography in the Mediterranean. *Taxon* **54:** 877-885.

Popov SV, Shcherba IG, Ilyina LB, Nevesskaya LA, Paramonova NP, Khondkarian SO, Magyar I. 2006. Late Miocene to Pliocene palaeogeography of the Paratethys and its relation to the Mediterranean. *Palaeogeography, Palaeoclimatology, Palaeoecology* **238:** 91-106.

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

Pritchard JK, Wen W. 2004. Documentation for STRUCTURE software version 2. Department of Human Genetics. University of Chicago, Chicago, IL. Available from: http://pritch.bsd.uchicago.edu.

Prunier J, Pelgas B, Gagnon F, Desponts M, Isabel N, Beaulieu J, Bousquet J. 2013. The genomic architecture and association genetics of adaptive characters using a candidate SNP approach in boreal black spruce. *BMC Genomics* **14:** 368.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rannala B, Mountain JL. 1997. Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences USA* **94:** 9197-9201.

Rodríguez-Sánchez F, Guzmán B, Valido A, Vargas P, Arroyo J. 2009. Late Neogene history of the laurel tree (*Laurus* L., Lauraceae) based on phylogeographical analyses of Mediterranean and Macaronesian populations. *Journal of Biogeography* **36:** 1270-1281.

Rousset F. 2008. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8:** 103-106.

Sanz R, Pulido F, Nogués-Bravo D. 2009. Predicting mechanisms across scales: amplified effects of abiotic constraints on the recruitment of yew *Taxus baccata*. *Ecography* **32:** 993-1000.

Savolainen O, Pyhäjärvi T, Knürr T. 2007. Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics* **38:** 595-619.

Schirone B, Caetano-Ferreira R, Vessella F, Schirone A, Piredda R, Simeone MC. 2010. *Taxus baccata* in the Azores: a relict form at risk of imminent extinction. *Biodiversity and Conservation* **19:** 1547-1565.

Shah A, Li D-Z, Möller M, Gao L-M, Hollingsworth ML, Gibby M. 2008. Delimitation of *Taxus fuana* Nan Li & R.R. Mill (Taxaceae) based on morphological and molecular data. *Taxon* **57:** 211-222.

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* **92:** 142-166.

Stewart JR, Lister AM. 2001. Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution* **16:** 608-613.

Temunović M, Franjić J, Satović Z, Grgurev M, Frascaria-Lacoste N, Fernández-Manjarrés JF. 2012. Environmental heterogeneity explains the genetic structure of Continental and Mediterranean populations of *Fraxinus angustifolia* Vahl. *PLoS One* **7:** e42764.

Thomas PA, Polwart A. 2003. Biological flora of the British Isles. *Journal of Ecology* **91:** 489-524.

Thompson JD. 2005. *Plant evolution in the Mediterranean*. Oxford, UK: Oxford University Press.

Turner C. 2000. The Eemian interglacial in the North European plain and adjacent areas. *Netherlands Journal of Geosciences* **79:** 217-231.

Tzedakis PC, Emerson BC, Hewitt GM. 2013. Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends in Ecology & Evolution* **28:** 696-704.

Wang IJ. 2013. Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution* **67:** 3403-3411.

Willis KJ, van Andel TH. 2004. Trees or no trees? The environments ofcentral and eastern Europe during the Last Glaciation. *Quaternary Science Reviews* **23:** 2369-2387.

Willyard A, Syring J, Gernandt DS, Liston A, Cronn R. 2007. Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. *Molecular Biology and Evolution* **24:** 90-101.

Wright S. 1931. Evolution in Mendelian populations. *Genetics* **16:** 97-159.

1
2
3
4

Supporting Information

Fig. S1 Geographical location of the twelve different sets of populations used for ABC simulations.

Fig. S2 Pre-evaluation of scenarios and prior distributions.

Fig. S3 Geographical distribution of the chloroplast haplotypes detected in the *trnS–trnQ* and *trnL–trnF* intergenic spacers.

Fig. S4 Summary of the clustering results using TESS for K=2, and STRUCTURE for K=3 and K=4.

Fig. S5 Prior and posterior distributions in the ABC analysis.

Fig. S6 Principal Component Analysis (PCA) plot of environmental variables for the present time described in Table S4.

Fig. S7 MAXENT predicted suitability for *Taxus baccata* at the whole range for present (PRE) and past (LGM, LIG) climatic conditions.

Fig. S8 Relationship between sex-ratio and temperature.

Table S1 Location, sample size (N), genetic diversity (A_R, H_E, H_O) and genetic cluster assignment of the studied populations. Retained populations for correlations with past climate are indicated (R).

Table S2 Sampled populations and polymorphic sites for the *trnS–trnQ* intergenic spacer.

Table S3 Sampled populations and polymorphic sites for the *trnL–trnF* intergenic spacer.

Table S4 Bioclimatic variables and standardized loadings for the two first axes of the PCA analysis (present climate).

Table S5 Pairwise F_{ST} values corrected for the possible presence of null alleles using the program FREENA.

Table S6 Analysis of molecular variance (AMOVA).

Notes S1 Details and results of model checking and confidence in scenario choice.

Notes S2 Species distribution models and correlations between genetic distance (F_{ST}) and environmental variables obtained using the "BIOCLIM" algorithm implemented in DIVA-GIS v.7.5.

Table 1 Demographic ABC models for *Taxus baccata*.

Simulation	Supported scenario (*PP*)	I (10^4)	E (10^4)	A (10^4)	W (10^4)	NA (10^5)	ra	t1	t2	t3
Considering two gene pools (Western, Eastern)										
sim1_500	A (>0.9)	-	7.87	0.78	4.43	8.13	0.589	0.071	2.110	-
	[0.9663,0.9895]		(4.34-9.81)	(0.18-4.25)	(2.1-7.13)	(2.77-9.83)	(0.204-0.867)	(0.016-0.281)	(0.444-7.950)	
sim2_500	A (>0.9)	-	6.39	3.09	4.61	6.42	0.431	0.164	2.480	-
	[0.9165,0.9681]		(2.74-9.53)	(0.87-8.19)	(2.12-7.68)	(1.28-9.59)	(0.108-0.789)	(0.037-0.644)	(0.516-8.380)	
sim3_500	A (~0.6)	-	6.43	4.26	2.92	3.19	0.207	0.195	1.870	-
	[0.5135,0.7619]		(2.83-9.51)	(1.33-8.89)	(1.01-6.90)	(0.33-8.47)	(0.032-0.661)	(0.041-0.907)	(0.365-7.960)	
sim4_500	B (~0.6)	-	6.65	3.99	4.21	6.92	-	0.313	1.870	-
	[0.5213,0.7602]		(2.89-9.64)	(1.87-6.80)	(1.77-7.67)	(2.00-9.72)		(0.079-1.150)	(0.408-7.660)	
sim5_500	B (~0.7)	-	3.31	4.71	4.54	4.77	-	0.573	3.200	-
	[0.6357,0.8293]		(0.96-8.38)	(1.93-8.60)	(1.83-8.33)	(0.84-9.19)		(0.145-2.070)	(0.736-8.620)	
sim6_500	A (>0.9)	-	6.08	0.62	4.15	0.51	0.414	0.062	0.841	-
	[0.8941,0.9701]		(2.67-9.40)	(0.14-3.93)	(1.47-8.36)	(0.04-3.50)	(0.070-0.871)	(0.012-0.316)	(0.151-5.360)	
sim7_500	A (>0.9)	-	7.25	0.86	3.58	4.34	0.507	0.069	1.800	-
	[0.9515,0.9865]		(3.58-9.67)	(0.20-4.86)	(1.09-7.28)	(0.57-9.02)	(0.112-0.880)	(0.015-0.346)	(0.348-7.560)	
sim8_500	B (>0.9)	-	7.39	2.60	6.63	5.95	-	0.480	1.730	-
	[0.8454,0.9581]		(3.65-9.76)	(1.22-4.58)	(3.40-9.33)	(1.42-9.57)		(0.138-1.410)	(0.385-7.350)	
sim9_500	A (~0.8)	-	5.53	4.83	3.71	7.22	0.285	0.288	1.940	-
	[0.6901,0.8630]		(2.11-9.40)	(1.61-9.00)	(1.57-6.60)	(1.95-9.76)	(0.064-0.622)	(0.065-1.060)	(0.376-7.900)	
sim10_500	A (>0.9)	-	6.72	3.37	5.08	5.78	0.374	0.124	2.270	-
	[0.9301,0.9745]		(3.14-9.63)	(0.92-8.21)	(2.38-8.18)	(0.93-9.48)	(0.088-0.756)	(0.026-0.500)	(0.479-8.160)	
Considering three gene pools (Western, Eastern, Iran)										
sim1_700	C (>0.9)	8.06	4.30	3.95	2.96	3.86	0.405	0.193	2.690	8.190
	[0.8746,0.9537]	(5.04-9.78)	(1.74-8.45)	(1.36-8.61)	(1.33-5.85)	(0.36-9.35)	(0.065-0.859)	(0.044-0.826)	(0.732-8.270)	(1.640-20.40)
sim2_700	C (>0.9)	7.59	4.00	2.96	3.09	7.42	0.529	0.202	1.790	4.030
	[0.8837,0.9585]	(4.24-9.74)	(1.69-8.00)	(0.90-7.77)	(1.56-5.27)	(2.50-9.76)	(0.127-0.880)	(0.054-0.630)	(0.586-5.070)	(1.050-9.150)

PP=posterior probability and [95% confidence intervals]; I=current effective population size of the *Iran* gene pool; E=current effective population size of the *Eastern* gene pool; A=current effective population size of the *Admixed* samples; W=current effective population size of the *Western* gene pool; NA=ancestral effective population size; ra=admixture rate; t1, t2 and t3 are estimated times of the different events depicted in Fig. 2 (in Myr). Values are medians (5 and 95% quartiles) of posterior distributions.

Table 2 Partial Mantel (PM) correlation (r) and Multiple Matrix Regression (MMRR) coefficients (b) between genetic distance (F_{ST}) and environmental variables for the present time time (PRE, ~1950-2000). Analyses were conducted considering the 195 populations of *Taxus baccata* with $N \geq 8$ (Whole range), and only populations within *Western* (153 sampling sites) and *Eastern* (64 sampling sites) gene pools.

		PRE		
		MMRR		PM
		$b_{Geo\text{-}PRE}$	$b_{Env\text{-}PRE}$	$r_{Env\text{-}PRE}$
Whole range	F_{ST} ~ PC1/Geo	0.305***	**0.154*** **	**0.157** **
	F_{ST} ~ PC2/Geo	0.355***	-0.043ns	-0.046ns
	F_{ST} ~ BIO1/Geo	0.328***	**0.161*** **	**0.171*** **
	F_{ST} ~ BIO6/Geo	0.289***	**0.145*** **	**0.141*** **
Eastern pool	F_{ST} ~ PC1/Geo	0.406***	0.037ns	0.040ns
	F_{ST} ~ PC2/Geo	0.457***	-0.143*	-0.148ns
	F_{ST} ~ BIO4/Geo	0.325***	**0.150***	**0.136***
Western pool	F_{ST} ~ PC1/Geo	0.098**	**0.175*** **	**0.174** **
	F_{ST} ~ PC2/Geo	0.126**	-0.002ns	-0.002ns
	F_{ST} ~ BIO1/Geo	0.108**	**0.218*** **	**0.219** **
	F_{ST} ~ BIO6/Geo	0.096*	**0.165*** **	**0.163** **

Variables accounting for PC1 (BIO1, BIO4, BIO9, BIO10, B11, BIO18) and PC2 (BIO12, BIO13, BIO16, BIO19) are described in Table S4. BIO1=Annual mean temperature. BIO4=Temperature seasonality. BIO6=Min temperature of the coldest month.

*** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$, ns=not significant.
Positive significant tests for both Multiple Matrix Regressions and Partial Mantel tests are in bold.

Table 3 Partial Mantel (PM) correlation (r) and Multiple Matrix Regression (MMRR) coefficients (b) between genetic distance (F_{ST}) and environmental variables for the last glacial maximum (LGM, ~21,000 yrs BP) and the last interglacial (LIG, ~120,000-140,000 yrs BP). The number of *Taxus baccata* populations retained for the analyses (i.e., with suitability values of MAXENT predicted distributions above the maximum training sensitivity plus specificity threshold) are indicated in brackets behind each period considered, and are specified in Table S1.

	LGM-MIROC (102)			LGM-CCSM (123)			LIG (94)		
	MMRR		PM	MMRR		PM	MMRR		PM
	$b_{Geo\text{-}MIROC}$	$b_{Env\text{-}MIROC}$	$r_{Env\text{-}MIROC}$	$b_{Geo\text{-}CCSM}$	$b_{Env\text{-}CCSM}$	$r_{Env\text{-}CCSM}$	$b_{Geo\text{-}LIG}$	$b_{Env\text{-}LIG}$	$r_{Env\text{-}LIG}$
F_{ST} ~ PC1/Geo	0.166**	-0.088*	-0.086ns	0.317***	-0.128**	-0.132ns	0.343***	0.056*	0.058 ns
F_{ST} ~ PC2/Geo	0.130**	0.077ns	0.077ns	0.285***	0.069ns	0.072ns	0.358***	0.015ns	0.016ns
F_{ST} ~ BIO1/Geo	0.164**	-0.062ns	-0.057ns	0.315***	-0.096*	-0.098ns	0.357***	**0.115***	**0.104***
F_{ST} ~ BIO2/Geo	0.137**	0.031ns	0.031ns	0.306***	-0.068ns	-0.070ns	0.318***	**0.170*****	**0.177****
F_{ST} ~ BIO4/Geo	0.219**	-0.199***	-0.185ns	0.314***	-0.120***	-0.123ns	0.490***	-0.235***	-0.209ns
F_{ST} ~ BIO6/Geo	0.142**	-0.005ns	-0.004ns	0.312***	-0.065ns	-0.065ns	0.303***	**0.106****	**0.096***

Variables accounting for PC1 were BIO1, BIO5, BIO6, BIO9, BIO10, B11 for LGM-MIROC, BIO1, BIO2, BIO5, BIO6, BIO8, BIO9, BIO10, B11 for LGM-CCSM, and BIO1, BIO2, BIO4, BIO6, BIO9, B11, BIO18 for LIG. Variables accounting for PC2 were the same for all periods considered, and the same as for PRE (BIO12, BIO13, BIO16, BIO19; Table S4). BIO1=Annual mean temperature. BIO2=Mean diurnal range (mean of monthly (max temp - min temp)). BIO4=Temperature seasonality. BIO6=Min temperature of the coldest month.

*** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$, ns=not significant.
Positive significant tests for both Multiple Matrix Regressions and Partial Mantel tests are in bold.

Figure Legends

Fig. 1 Location of the 238 populations of *Taxus baccata* included in this study (red circles: populations with $N \geq 8$, white circles: populations with $N<8$). The geographical distribution of the species is shown in blue (kindly provided by EUFORGEN, the European Forest Genetic Resources programme, www.euforgen.org). A feminine cone of the species is shown in the inset.

Fig. 2 Demographic scenarios used for *Taxus baccata* in the ABC analyses. Considering two gene pools (*Western*, *Eastern*): A) populations in central Europe, Italy and the Mediterranean islands were originated from "secondary contact" between those from *Eastern* and *Western* origin; B) "colonization" from the Eastern territories to the Mediterranean area. Considering three gene pools (*Western*, *Eastern*, *Iran*): C) "colonization" of eastern Europe from *Iran*, separation of European samples into two genetic pools (*Eastern*, *Western*), and subsequent "secondary contact" of both pools in central Europe, Italy and the Mediterranean islands; D) as in scenario B, "colonization" from the Eastern territories to the Mediterranean area, but considering three groups of populations. NA=ancestral effective population size; t1, t2 and t3=divergence times for the depicted events.

Fig. 3 Best number of genetic clusters (K) obtained for *Taxus baccata* using STRUCTURE (K=2) and TESS (K=3). The eight populations with significant linkage among loci are excluded. Pie charts show averaged values of the different runs for the proportion of membership to each genetic cluster.

Fig. 4 Distribution of genetic diversity (A_R: allelic richness; H_E: unbiased expected heterozygosity) in *Taxus baccata*. Values for A_R and H_E are indicated by the circle size (○>○>○) and the colour gradient (red>orange>yellow>green), respectively.

Fig. 5 MAXENT predicted suitability for *Western* and *Eastern* gene pools of *Taxus baccata* during three time periods: LIG=Last interglacial (~120,000-140,000 yrs BP), LGM-CCSM and LGM-MIROC=Last Glacial Maximum (~21,000 yrs BP), PRE=present conditions (~1950-2000). Darker colours indicate higher probabilities of suitable climatic conditions. Not suitable areas and those with logistic output values

below the maximum training sensitivity plus specificity (MTSS) threshold are indicated in grey.