Scots pine – panmixia and the elusive signal of genetic adaptation

Jade Bruxaux[1†], Wei Zhao[1†], David Hall[1,2†], Alexandru Lucian Curtu[3], Piotr Androsiuk[4], Andreas D. Drouzas[5], Oliver Gailing[6], Heino Konrad[7], Alexis R. Sullivan[1], Vladimir Semerikov[8], Xiao-Ru Wang[1*]

[1]Department of Ecology and Environmental Science, Umeå Plant Science Center, Umeå University, Umeå, Sweden.

[2]Forestry Research Institute of Sweden (Skogforsk), Sävar, Sweden.

[3]Department of Silviculture, Transilvania University of Braşov, Braşov, Romania.

[4]Department of Plant Physiology, Genetics and Biotechnology, Faculty of Biology and Biotechnology, University of Warmia and Mazury in Olsztyn, Olsztyn, Poland.

[5]Laboratory of Systematic Botany and Phytogeography, School of Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece.

[6]Department of Forest Genetics and Forest Tree Breeding, University of Göttingen, Göttingen, Germany.

[7]Austrian Research Centre for Forests (BFW), Department of Forest Biodiversity and Nature Conservation, Unit of Ecological Genetics, Vienna, Austria.

[8]Institute of Plant and Animal Ecology, Ural Division of Russian Academy of Sciences, Ekaterinburg, Russia.

[†]These authors contributed equally to the work.

[*]Correspondence to Xiao-Ru Wang, xiao-ru.wang@umu.se

Summary

- Scots pine is the foundation species of diverse forested ecosystems across Eurasia and displays remarkable ecological breadth, occurring in environments ranging from temperate rainforests to arid tundra margins. Such expansive distributions can be favored by various demographic and adaptive processes and the interaction between them.
- To understand the impact of neutral and selective forces on genetic structure in Scots pine, we conducted range-wide population genetic analyses on 2,321 trees from 202 populations using genotyping-by-sequencing, reconstructed the recent demography of the species, and examined signals of genetic adaptation.
- We found a high and uniform genetic diversity across the entire range, with no indication of increased genetic load in expending populations and minor impact of the last glacial maximum on historical population sizes. Genetic-environmental associations identified only a handful of SNPs significantly linked to environmental gradients.
- The results suggest that extensive gene flow is predominantly responsible for the observed genetic patterns in Scots pine. The apparent missing signal of genetic adaptation is likely attributed to the intricate genetic architecture controlling adaptation to multi-dimensional environments. The panmixia metapopulation of Scots pine offers a good study system for further exploration into how genetic adaptation and plasticity evolve under gene flow and changing environment.

Keywords

Conifer; demography; genetic diversity; gene flow, genetic-environmental association, population structure; *Pinus sylvestris.*

Introduction

Life history traits, demographic events, and selection shape the amount and distribution of genetic diversity across a species range (Ellegren & Galtier, 2016). Wind-pollinated outcrossing species are for example generally expected to have higher genetic diversity and weaker genetic structure compared to inbreeding species. In a heterogeneous fitness landscape, this high genetic diversity would provide a greater adaptive potential and promote divergence under environment-specific selection (Hamrick & Godt, 1996; Savolainen *et al.*, 2013; Gamba & Muchhala, 2020). Consecutive range expansions after contraction may carry only a portion of the diversity from the source population and produce differentiation via genetic drift (Hewitt, 1996; Orsini *et al.*, 2013; Wang & Bradburd, 2014; Zhao *et al.*, 2020). These processes could potentially constrain a species ability to track, tolerate, and adapt to changing environments through their impact on standing adaptive genetic diversity or on gene flow between populations. Understanding the legacy effect of these processes is particularly important for boreal forest trees, which are subjected to a strong ongoing climate change (Davis & Shaw, 2001; Koenigk *et al.*, 2020).

Scots pine (*Pinus sylvestris* L.) is the foundation species of diverse forested ecosystems across Eurasia. It has the widest distribution of any pine species, with a range spanning from the Iberian and Balkan Peninsulas north to the Barents Sea and from the British Isles east to the Russian Pacific coast (Fig. **1**; Carlisle & Brown, 1968). It has an exceptionally wide niche, tolerating annual average temperatures ranging from -3°C to 16°C, and annual precipitation between 400 and 2,900 mm (Houston Durrant *et al.*, 2016). It can be found from the sea level in its most northern populations and up to 2,600 m above the sea level in the Caucasus (Houston Durrant *et al.*, 2016). It is a light demanding species able to grow on very nutrient poor sites (Houston Durrant *et al.*, 2016). The variable habitat types that Scots pine occupies appear to have promoted distinct local adaptation as reflected by characteristic growth forms, ecotypes and clines in physiological traits, with up to 150 morphological varieties described (Carlisle & Brown, 1968; Rehfeldt *et al.*, 2002; Tóth *et al.*, 2017a). Phenotypic differences on traits such as bud flush and set, height increment, growth rate, and frost hardiness are extensive as demonstrated by provenance trials (Hurme *et al.*, 1997; Andersson & Fedorkov, 2004; Hall *et al.*, 2021). Variations in these traits have been correlated with geographic parameters such as latitude or longitude, and with environmental parameters such as temperature, precipitation and the length of the growing season. Clinal phenotypic variation is a response to spatially varying selection along environmental gradients, which is expected to

generate gene frequency clines at associated loci (Savolainen *et al.*, 2007). This phenomenon can also arise from plastic response (Vitasse *et al.*, 2010; McLean *et al.*, 2014), or a combination of both genetic and plastic mechanisms (Corl *et al.*, 2018).

The current distribution of Scots pine is believed to have been shaped by migration out of several scattered refugia that survived the last glacial maximum (LGM) in southern Europe, Turkey and the Russian Plain (Willis *et al.*, 2000; Tóth *et al.*, 2017a). Mitochondrial polymorphisms suggest that Scots pine expanded into Fennoscandia from west-central Europe and the Russian Plain 10,000 – 7,000 years ago as they tracked receding ice sheets from ice age refugia (Naydenov *et al.*, 2007; Pyhäjärvi *et al.*, 2008). However, several recent discoveries question the local extinction of the species during that period and suggest the existence of northern refugia. A tree dated at 11,700 before present (BP) has for example been discovered in central Sweden (Kullman, 2002), which would imply an extremely fast migration rate from the southern Europe refugia if the species had locally disappeared. A vegetation model identified potential *P. sylvestris* refugia during the LGM up to 50°N (Cheddadi *et al.*, 2006), based on the European Pollen database and charcoal records showing a continuous presence of conifers during 32-16,000 BP in central and eastern Europe (Willis & van Andel, 2004). Finally, the observed high chloroplast diversity at intermediate latitudes for 22 European trees and shrubs, including Scots pine, are difficult to explain with only a few small refugia in southern Europe (Petit *et al.*, 2003).

Rapid colonization from few isolated refugia is expected to leave distinct genetic signature in the extant populations such as declining genetic diversity along migration routes and genetic lineages that track their recent ancestry (Hewitt, 2000; Excoffier *et al.*, 2009). In such a scenario, we would expect lower genetic diversity in the northern populations of Scots pine and higher differentiation of them from the refugial populations. However, if the species survived in many micro-refugia scattered across the current distribution, we should observe little decline of the diversity in the north and weak differentiation between populations. Testing these hypotheses requires comprehensive coverage of the species distribution and sufficient genetic information. However, genome-wide patterns of genetic diversity across the species range, as well as the impact of landscapes and environments on population differentiation, remain scarcely studied for Scots pine (but see Sannikov & Petrova, 2012; Tyrmi *et al.*, 2020), hindering the recovery of its evolutionary history and genetic signature of local adaptation.

In this study, we conducted genotyping-by-sequencing on 2,321 Scots pine individuals sampled from more than 200 populations across its natural range. With this extensive sampling, we aimed to assess: 1) the composition and distribution of genetic diversity in Scots pine at

regional and range-wide scales, and the impact of geography, ecological conditions and gene flow on the spatial patterns, 2) the recent demographic history of the species, and whether expanding populations harbor high genetic load; and 3) the genetic and environmental associations, and the climatic drivers of population differentiation and local adaptation. These analyses enhance our understanding of how diversity correlates with geography, demography, and climate adaptation, and of the adaptive potential of the species in response to current and future environmental changes.

Material and methods

A more complete methods section is available in Supplementary Methods.

Sampling and genotyping-by-sequencing (GBS) library preparation

We collected 2895 samples from 211 natural stands covering the entire geographic distribution of *P. sylvestris*. The samples were either needles, buds or seedlings grown from seeds, collected in native forests to the best of our knowledge. Because the Scots pine genome is large (ca. 23Gbp; Fuchs *et al.*, 2008), we used a reduced representation sequencing method called GBS. The library preparation followed the protocol of Pan *et al.* (2015), using a *Pst*I-HF enzyme to digest the DNA. After fragment size selection, paired-end sequencing (2x150 bp) was performed on Illumina HiSeq X Ten (Novogene, UK).

Genotype likelihoods

After preliminary analyses detailed in Supplementary Methods, we retained 2,321 unrelated individuals from 202 stands (Fig. **1a**, Fig. **S1**, Table S1). Clean reads were mapped on the *P. taeda* genome v.1.01 (Neale *et al.*, 2014; Zimin *et al.*, 2014) using the Burrows-Wheeler Aligner v. 0.7.17 (Li, 2013). Instead of calling SNPs, we used ANGSD v. 0.935 (Korneliussen *et al.*, 2014) to estimate genotype likelihoods, *i.e.* the probability of each genotype at each site and for each individual. We set a maximal depth of 99.5% of the highest depth measured (5022.8x), a minimal depth of 5, a Phred-scaled base quality of 20 and a mapping quality of 40 for an individual site to be considered. We removed the sites with >40% missing data, more than two alleles and an observed heterozygosity higher than 50%.

The samples were sequenced in 13 different libraries, thus slight shift in size selection can cause a certain degree of non-overlap of the sequenced sites between libraries, creating

similarity by sequence library instead of geographical origin of the samples. To remove this batch or library effect, we identified the SNPs grouping subsets of samples by library instead of country or region using PCAngsd v. 1.01 (Meisner & Albrechtsen, 2018) and removed them. We re-ran ANGSD using the final SNP list and all the invariant sites for nucleotide diversity and demographic analyses. We produced another SNP dataset with a minimal allele frequency (MAF) of 0.05 for all the other analyses. We checked that the final SNPs were independent by estimating the linkage disequilibrium between each pair of SNPs using ngsLD v. 1.2.0 (Fox *et al.*, 2019).

Genetic diversity and population structure

To describe the level and structure of genetic diversity at regional and range-wide scales, we first combined the stands having less than six individuals with their closest neighbors to reach at least six individuals per population, the minimal number to obtain reliable diversity estimates (Nazareno *et al.*, 2017). Three stands from central Russia that were too distant (>375km; Table S1) to any other sampled stands were left alone. This way, we obtained 123 larger populations, covering the entire distribution (Fig. **1a**, Fig. **S1**). We estimated for each large population the per-site inbreeding coefficient (F_{IS}), observed (H_o) and expected heterozygosity (H_e) with ANGSD on the MAF 0.05 dataset. We estimated the pairwise nucleotide diversity (π) and Tajima's D statistic (Tajima, 1989) from the folded site frequency spectrum (SFS) based on the full dataset. The 95% confidence interval (CI) of Tajima's D was estimated by subsampling the data with replacement to produce 100 SFS. The nucleotide diversities at 0-fold sites (π_0, where all mutations would be non-synonymous) and 4-fold sites (π_4, where all mutations would be synonymous) were based on the gene annotations of the *P. taeda* genome.

To characterize the spatial organization of the genetic diversity, we first performed a PCA on the MAF filtered dataset with PCAngsd including all samples. As PCA only performs linear dimension reduction, we added a Uniform Manifold Approximation and Projection (UMAP) analysis which implements non-linear dimension reduction using UMAP v. 0.5.3 (McInnes *et al.*, 2018). We then examined the population structure using NGSadmix v. 32 (Skotte *et al.*, 2013) with a potential number of clusters K ranging from 1 to 20, with 50 replicates for each analysis, and a maximum of 200,000 iterations. We tried to identify the optimal K value representing the data using both the Evanno and the logarithm method (Pritchard *et al.*, 2000; Evanno *et al.*, 2005).

To describe geographic patterns of genetic differentiation (F_{ST}) between pairs of large populations, we first estimated F_{ST} using ANGSD. We then performed Mantel tests to evaluate

the spatial autocorrelation of the results, comparing geographic *vs* genetic distance between populations (normalized F_{ST}) using the R package vegan v. 2.5-6 (Oksanen *et al.*, 2019) and 9999 permutations, genetic *vs* environmental distance (Mahalanobis distance calculated within the RDA; see below), and geographic *vs* environmental distance.

To account for the impact of isolation by distance (IBD; Wright, 1943) on population differentiation, we examined population structure using tess3r which considers IBD in defining genetic clusters (Caye *et al.*, 2016). We ran the tess3r analysis for *K* ranging from 1 to 20, with 20 replicates for each value of *K* and a maximum of 200 iterations.

Finally, to discern spatial patterns of gene flow, we estimated the effective migration surfaces with the python package feems (Marcus *et al.*, 2021). This method is based on a stepping-stone model and identifies migration rates between populations that deviate from what would be expected under pure IBD, thereby locating gene flow barriers or corridors. We performed a leave-one-out cross-validation for lambda values (tuning parameter) varying from 10^{-6} to 10^{2}, with 20 values tested.

Demographic history

As historic events can have a significant impact on population genetic diversity and differentiation, we estimated demographic changes of the three most divergent populations in Scots pine (i.e. China, Spain, and northern Norway) using two different approaches. The first, Stairway Plot v. 2 (Liu & Fu, 2015; Liu & Fu, 2020), performs multi-epoch coalescent inference of population size changes through time. It is suitable for demographic analyses where no previous knowledge is available but can only analyze populations independently. The second, fastsimcoal2 v. 2.6 (Excoffier *et al.*, 2013; Excoffier *et al.*, 2021) implements multi-population demographic models that take divergence and migration among populations into consideration.

We produced folded SFS with ANGSD on the full dataset (including invariant sites) for these three populations and 2D-SFS for each pair of populations. We ran Stairway Plot for each of the three populations separately and for the entire species (2,321 individuals) under the default parameters. We used an average generation time of 20 (Pyhäjärvi *et al.*, 2020) and 50 years (Willyard *et al.*, 2007; Tóth *et al.*, 2019) and a mutation rate of 7×10^{-10} mutations per site and year (Willyard *et al.*, 2007). We considered the entire allele frequency spectrum with or without singletons.

We tested 20 demographic scenarios with fastsimcoal2 excluding singletons (Fig. **S2**). These 20 models cover most of plausible population history scenarios. Each model was run 50 times, with 100,000 coalescent simulations and 40 expectation-conditional maximization cycles

for the calculation of the global maximum-likelihood. The best-fitting model was selected based on the maximum value of likelihood over the 50 independent runs of each model and their Akaike's weight of evidence (Akaike, 1987). Parameter confidence intervals (95% CI) of the best model were obtained by running 100 parametric bootstraps, with 50 independent runs in each bootstrap. We used a mutation rate of 7×10^{-10} mutations per site and year and a generation time of 50 years.

To estimate the potential fitness impact of population expansion (see Results), we first compared the efficacy of purifying selection as measured by π_0/π_4 ratio with the Tajima's D value for each large population. If demographic expansion is correlated with an increase in genetic load (higher frequency of deleterious mutation), we should observe a negative correlation between the two parameters. We also looked at the gene effects of the SNPs by classifying them using snpEff v. 5.2 (Cingolani *et al.*, 2012). We then estimated the additive and recessive genetic load at the individual level as in de Pedro *et al.* (2021) and compared the average values between large populations.

Genotype and environment association (GEA)

To explore potential signals of genetic adaptation driven by local environmental conditions, we extracted 74 soil, climate, light conditions, and species composition variables from several databases (Table S2) to describe the environment of each of the 123 large populations. Subsequently, we employed a gradient forest analysis performed with the R-package gradientForest v. 0.1-37 (GF; Ellis *et al.*, 2012) to rank them by importance, a random forest concept that evaluates the significance of a predictor for the model. The GF models were validated by a permutation test through 10 different runs using shuffled environmental data. The variables with a correlation coefficient $\geq |0.75|$ (Spearman's R) to the top-ranked variable were removed. This was done with every subsequent top ranked variable. The remaining variables were then used as input for the subsequent analyses.

We performed redundancy analyses (RDA; Legendre & Legendre, 2012; Rellstab *et al.*, 2015) to evaluate the impacts of three explanatory matrices on allele frequencies across populations. The first matrix was based on the distance between populations through principal coordinates of neighborhood matrix (PCNM) that converts distances to a rectangular representation (Borcard & Legendre, 2002). The second was based on the environmental variables extracted from the GF analysis, and the third was the shared coancestry between populations quantified from the first two UMAP+PCA axis. The first two matrices were first analyzed separately through a forward selection step with the 'ordistep' function in the vegan

package (Oksanen *et al.*, 2019). We ran a set of full and partial RDA models to quantify the impact of each of the three factors. Outliers in RDA were identified by comparing observed and predicted *p*-values distributions (modeled as logistic) of the number of SNPs scored.

We further conducted a BayPass analysis (Gautier, 2015) to identify significant SNPs associated with environmental variables. We first used the core model to establish the covariance of allele frequencies among populations that would be the result of their shared history. We detected outliers that showed strong differentiation among populations based on the observed log-normal distribution of the differentiation X^tX (*p*-value <0.001, see Fig. **5d**). These outliers are analogous to F_{ST} outliers. We then ran the auxiliary covariate model with spatial dependency on the markers using an Ising prior, b_{is} = 1, for each of the GF selected environmental variables independently, considering both outliers detected by the X^tX statistics and other non-differentiated SNPs. We considered a SNP to be correlated with an environmental variable when the standard deviation of the effect estimate did not cross zero ($|\beta|$-σ (β) > 0), corresponding to posterior inclusion probability ≥ 0.55. Finally, we identified the environmental variables that showed an association with at least one SNP and ran the previous model with all these variables simultaneously for final result. Significant SNPs identified by any GEA analysis were annotated based on the *P. taeda* genome annotations.

Results

Genotype dataset

The 2,321 samples that passed the quality filters averaged 2.42 million (M) mapped reads (0.33 - 49.56 M reads), 2.83 Mbp (million base pairs) with at least 5x depth (1.37 - 7.58 Mbp), and a read depth of 85.26x (17.38 - 953.61x). The final genotype likelihood dataset included 1,432,399 sites, of which 386,426 were SNPs. Filtering for MAF > 0.05 reduced the dataset to 25,830 independent SNPs. Among them, only 5,623 were in coding sequences, with 2,078 being synonymous mutation, 3,185 being non-synonymous mutation, and 96 being gain or loss of a stop codon.

Genetic diversity and population structure

We calculated the pairwise nucleotide diversity (π) among samples for each of the 123 large populations and found little variation among them with an average of 0.0062 differences per base pair (Table 1, Fig. **S3**). Across the whole distribution, π values, including π_0, π_4 and the

Table 1. Summary of the genetic diversity in Scots pine for each country/region and for the whole distribution. The number of populations (N_{pop}) and the number of individuals (N_{ind}) for each country or region are given in the two first columns. The overall pairwise nucleotide diversity (π), diversities at the 0-fold (π_0) and 4-fold (π_4) sites and Tajima's D were calculated on the full dataset including invariant sites. The inbreeding coefficient F_{IS}, the observed and expected heterozygosities (H_o and H_e) were based on the dataset filtered for MAF 0.05. Six individuals from three very distant Russian populations were not included in these analyses. The values presented are averages of the individual population results, weighted for the number of individuals per population. Tajima's D significance was tested by a one sample t-test on the 100 replicates and all results were highly significant (p-value < 10^{-3}). Results by large population are available in Table S1.

Region / country	**N_{pop}**	**N_{ind}**	**π**	**π_4**	**π_0**	**H_o**	**H_e**	**F_{IS}**	***Tajima's D***
Whole distribution		2315	0.00622	0.00890	0.00490	0.258	0.277	0.057	-1.297
Spain	3	70	0.00631	0.00900	0.00504	0.265	0.273	0.027	-1.369
Scotland	4	59	0.00638	0.00903	0.00504	0.284	0.292	0.020	-1.032
Western and Central Europe		526	0.00650	0.00928	0.00520	0.282	0.290	0.016	-1.191
France	25	271	0.00655	0.00941	0.00530	0.284	0.293	0.018	-1.158
Luxembourg	1	26	0.00640	0.00913	0.00505	0.258	0.268	0.033	-1.492
Germany	3	37	0.00663	0.00919	0.00519	0.278	0.287	0.019	-1.306
Austria	4	53	0.00643	0.00906	0.00507	0.285	0.292	0.017	-1.125
Slovenia	1	17	0.00641	0.00919	0.00502	0.278	0.280	0.007	-1.252
Slovakia	1	9	0.00634	0.00900	0.00509	0.310	0.300	-0.036	-0.976
Poland	4	68	0.00641	0.00912	0.00513	0.266	0.277	0.026	-1.325
Belarus	2	20	0.00621	0.00921	0.00495	0.299	0.296	-0.018	-0.892
Estonia	1	6	0.00633	0.00911	0.00506	0.346	0.325	-0.072	-0.776
Bulgaria	1	6	0.00632	0.00912	0.00494	0.353	0.329	-0.076	-0.798
Ukraine	1	13	0.00669	0.00949	0.00534	0.270	0.288	0.045	-1.401
Southeastern Europe		140	0.00644	0.00913	0.00513	0.267	0.275	0.024	-1.356
Greece	1	12	0.00709	0.00982	0.00564	0.324	0.304	-0.056	-1.302
Romania	3	60	0.00641	0.00916	0.00518	0.261	0.274	0.039	-1.377
Turkey	3	68	0.00635	0.00898	0.00500	0.263	0.271	0.026	-1.347
Eastern Europe and Asia		589	0.00610	0.00878	0.00481	0.268	0.283	0.045	-1.102
Russia	34	485	0.00613	0.00885	0.00483	0.266	0.285	0.051	-1.109
Kazakhstan	2	18	0.00612	0.00892	0.00487	0.317	0.304	-0.045	-0.819
Mongolia	1	13	0.00616	0.00892	0.00494	0.264	0.297	0.085	-1.013
China	3	73	0.00591	0.00828	0.00467	0.265	0.269	0.016	-1.139
Northern Europe		931	0.00610	0.00871	0.00473	0.235	0.265	0.099	-1.483
Norway	9	398	0.00607	0.00862	0.00466	0.236	0.265	0.100	-1.488
Sweden	10	351	0.00612	0.00880	0.00480	0.236	0.264	0.093	-1.505
Finland	6	182	0.00610	0.00871	0.00472	0.233	0.267	0.109	-1.428

π_0/π_4 ratio, decreased slightly with increasing latitude and longitude (Fig. **S4**). The average observed heterozygosity (H_o) was 0.258, and the expected heterozygosity (H_e) was 0.277 (Table 1, Table S1), with a small decrease at high latitudes, but no significant variation along longitude (Fig. **S4**). As anticipated for a highly outcrossing species, we recovered a very low inbreeding coefficient (F_{IS}) at most locations, averaging 0.059 and ranging from -0.088 to 0.2 (Table 1, Table S1). Slightly elevated F_{IS} values were observed in northern Norway (e.g., Alta, Vågå or Kirkesmoen) and in a few isolated populations (Black Sea or Massif Central in France), with a

general tendency for marginally higher values at high latitudes (Fig. **S4**). Tajima's D values were all highly negative (Fig. **S3**, Table 1, Table S1) and significantly different from 0 (one sample t-test; *p*-values < 10^{-3}). As these are genome-wide estimations, they can be interpreted as recent population expansion. While no correlation between Tajima's D with latitude was observed, a weak correlation with longitude was noted, with higher values in Asia (Fig. **S4**).

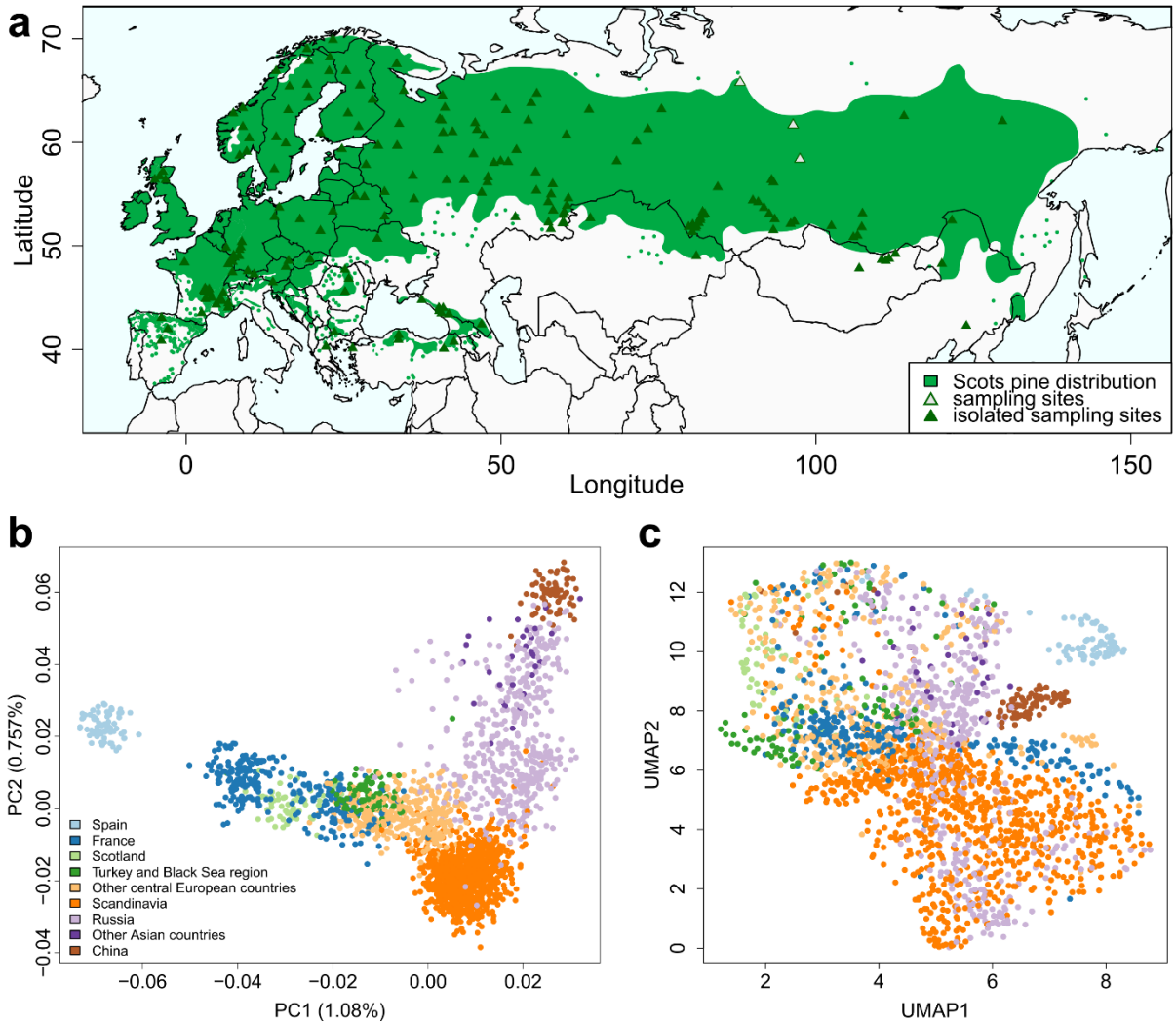


Figure 1: Sampling and genetic distribution in Scots pine. Panel **a**: Geographic distribution of the species with sampling sites. Green triangles represent sampling sites. Three populations indicated by grey triangles were used only in PCA and admixture analyses. Species distribution map from Caudullo *et al.* (2017). Panel **b**: Principal component analysis (PCA) with individuals colored by origin. The same figure colored by country is available in Fig. S4. Panel **c**: Uniform Manifold Approximation and Projection (UMAP) with individuals colored as in panel B.

On the PCA performed on the full collection (Fig. **1b**, Fig. **S5a**), we observed a pattern congruent with the geographic distribution: the first axis (PC1) followed a rough South-North

direction in west-central Europe, starting from Spain on the left and reaching Scandinavia on the right; the second axis (PC2) followed a more west-east direction for northern European, Russian, and Asian populations. The distribution was rather continuous, with only the Spanish populations being separated from the rest of the samples. The first two PC axes, however, explained less than 2% of the overall genetic variation, meaning that almost all the diversity is shared among all populations. The third and fourth axes, explaining less than 0.4% each, separated the Scottish and Turkish / Black Sea populations from the rest of the samples (Fig. **S5b**). The UMAP analysis confirms the very low level of structure in the data (Fig. **1c**), with only the Spanish and Chinese populations being slightly differentiated from the rest of the sampling, as well as one Romanian population (from Piatra Craiului National Park).

In the clustering analysis, we did not recover a clear optimal number of clusters (*K*): the Evanno method indicated that two groups explained the genetic diversity best, while the logarithm method indicated 20 groups (Fig. **S6**). At *K*=2, we observed a clear gradient between the western (Spanish) and eastern (Chinese) populations, corresponding to the two most distant ones (Fig. **2a**). At *K*=3, the Scandinavian populations formed a new cluster (Fig. **2b**). At *K*=4, a component for western Europe populations appeared. The Russian populations west of the Ural Mountains differentiated from the rest of the Asian populations if we considered five clusters. The Turkish and Black Sea populations split from the western Europe populations at *K*=6 (Fig. **2c**). One Romanian population (from Piatra Craiului National Park) stood out from nine clusters, and the Scottish populations were differentiated when we considered 10 groups. At *K*=12, the Caucasus populations separate from Asia Minor and no additional pattern emerged at *K* >12 (Fig. **S7**). Similarly, tess3r did not identify a specific number of clusters that best explained the genetic diversity in the population based on cross-validation analysis (Fig. **S8**), and the tess3r results (Fig. **S9**) were very similar to the ones obtained with NGSadmix.

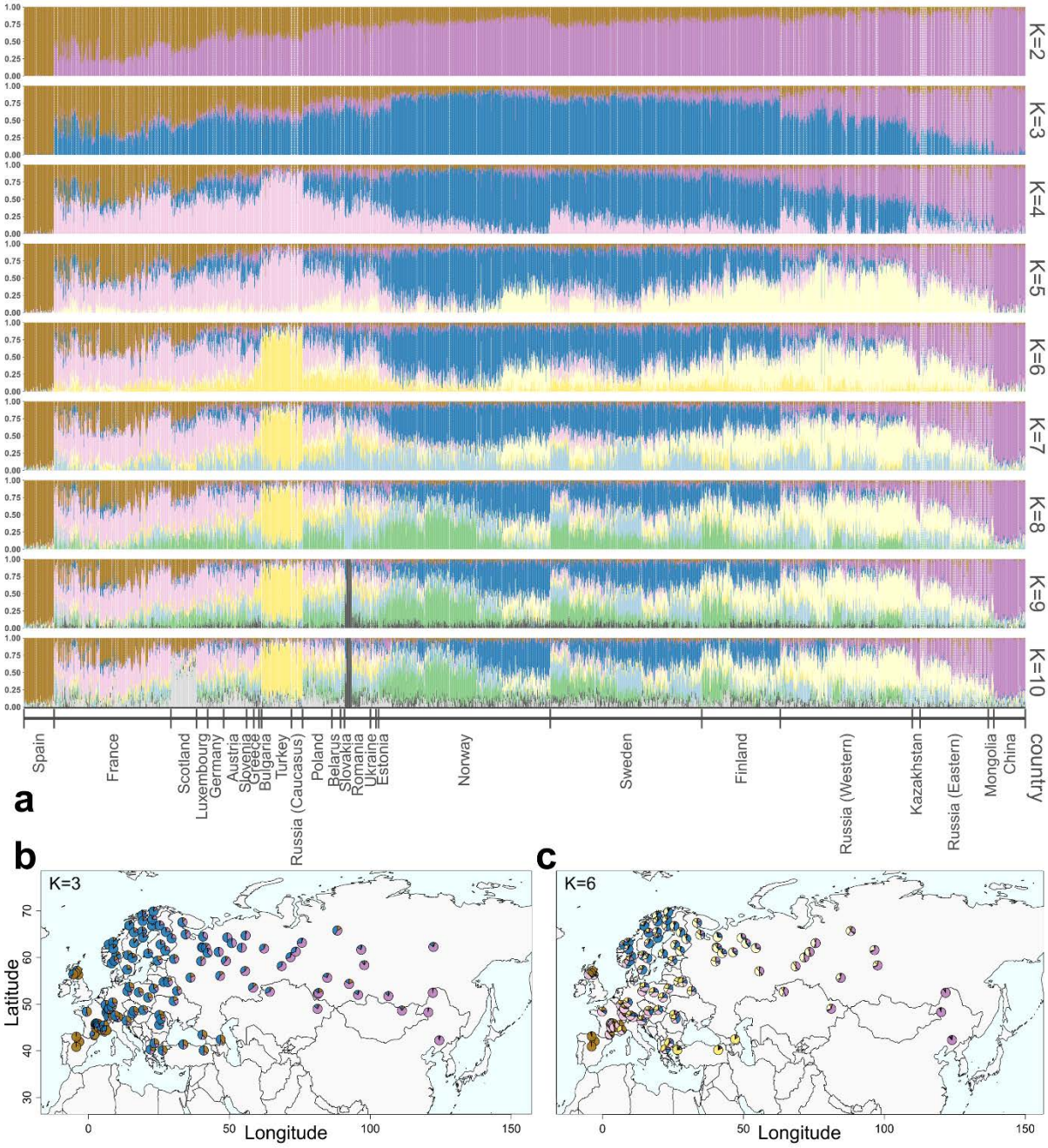


Figure 2: Genetic clustering in Scots pine. Panel **a**: Admixture inference showing the proportion of each cluster component for K = 2 - 10. Each color represents one genetic cluster. The populations are ordered by country following the PCA results: from Spain to Scandinavia then to China. In each country, the populations are ordered by longitude except for Scandinavia where they are ordered by latitude. For clustering patterns at K >10, see Fig. **S6**. Panel **b** and **c**: Cluster distribution on a map for K=3 and K=6. The colors follow the result in panel **a**. To reduce the number of pies and increase the readability, the populations represented here are the merged populations (Fig. **S1**).

The pairwise F_{ST} between populations ranged from 0.009 to 0.127, and the global F_{ST} among the 123 large populations averaged 0.048. In general, the genetic differentiation correlated strongly with the geographic distance separating the populations (Fig. **3a**; Mantel

test ρ: 0.586, *p*-value < 2.2×10^{-16}), showing a pattern of IBD. By contrast, the correlation of the genetic differentiation with the environmental distance was weak (Fig. **3b**; Mantel test ρ: 0.059, *p*-value: 4.5×10^{-7}), indicating no clear pattern of isolation by environment (IBE; Wang & Bradburd, 2014). The correlation between the geographic and environmental distances was also weak (Fig. **3c**; Mantel test ρ: 0.104, *p*-value < 2.2×10^{-16}), which means that they can be considered independent in the following analyses. In the feems analysis, we identified several regions with slightly reduced gene flow: the plains around the Pyrenees, Caucasus Mountains and Khingan Mountains, or the North Sea and Aegean Sea (Fig. **3d**).

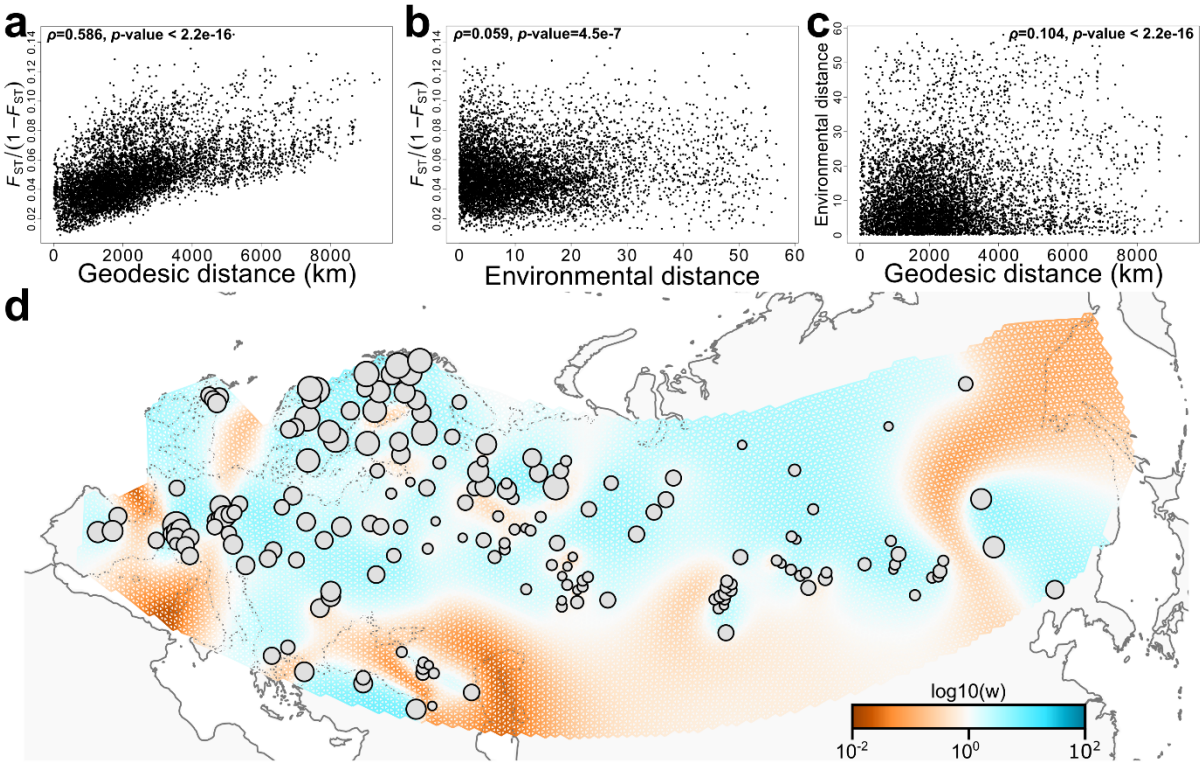


Figure 3: Spatial genetic patterns. Panel **a**: isolation by distance, with Mantel test on geographic *vs* genetic distance between populations. The genetic distance is presented by normalized F_{ST} between populations. Panel **b**: isolation by environment with Mantel test on environmental *vs* genetic distance. Panel **c**: Relationship between geographic and environmental distance. Panel **d**: Estimation of the effective migration surface between pairs of populations across the distribution range of Scots pine. Circles correspond to the sampled populations, with radius proportional to the sampling size. Brown areas correspond to a reduction of the gene flow compared to what would be expected given the distance separating populations, and blue areas correspond to an increase of the gene flow.

Demographic history

When excluding singletons and assuming a generation time of 50 years, we detected with the Stairway Plot 2 analysis an initial phase of population expansion starting around 6 Mya (million

years ago, Fig. **4a**, Fig. **S10**). The Spanish population was the first to expand, followed by the Chinese and Scandinavian populations, until they reached their contemporary sizes around 300 kya (thousands of years ago) for the Spanish population, 80 kya for the Chinese population, and 15 kya for the Scandinavian population. Both the Chinese and Scandinavian population sizes peaked at around 150 kya before reaching their contemporary sizes. For the entire species, we observed a sharp increase of the population size starting slightly before 1 Mya and then a rapid contraction at ca. 400 kya. The maximal population size was reached around 300 kya, with a decrease around 20 kya, corresponding to the LGM (Patton *et al.*, 2016), followed by a growth reaching 1.1 million individuals around 3,000 years ago. Results for a generation time of 20 years and for analyses with singletons are given in Fig. **S10**.

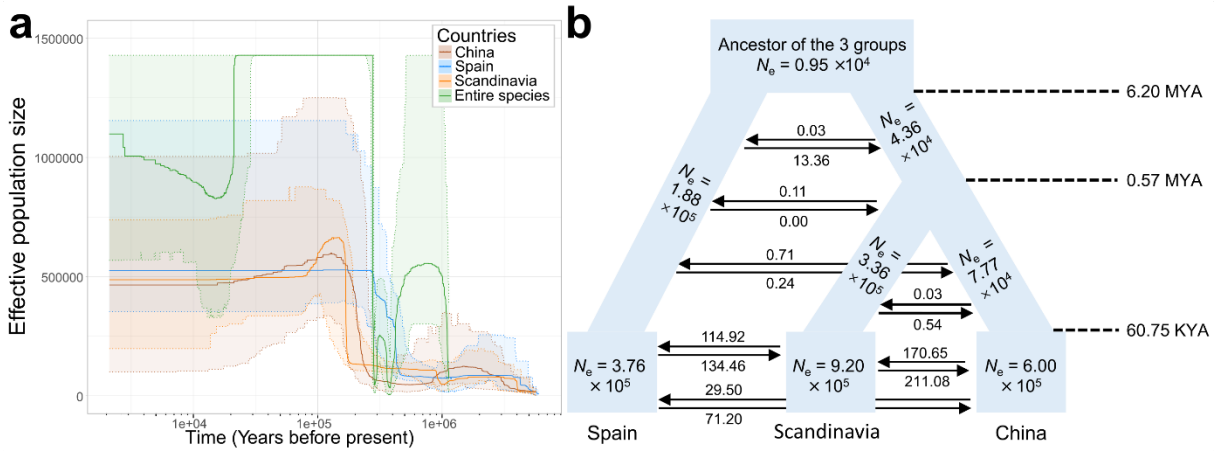


Figure 4: Population history of Scots pine. Panel **a**: Stairway Plot 2 analysis of the three main lineages and of the entire species. The plain lines represent the estimated median effective population sizes (N_e), and the shaded areas correspond to the 95% confidence interval. Panel **b**: fastsimcoal2 analysis. Each block represents a current or ancestral population with their estimated N_e. Arrows indicate the direction of gene flow with the estimated migration rate labeled above or below the arrow. The timing of the splitting into two groups and their size change are indicated in million years ago (MYA) or thousand years ago (KYA), respectively. The estimates were based on a mutation rate of 7×10^{-10} mutations /site /year and a generation time of 50 years, omitting singletons.

Among the 20 demographic models tested with fastsimcoal2, the first two best-fitting models (Model 7 and Model 3) were three-populations models with a pattern of isolation-with-migration (Fig. **S2**, Table S3). The expected SFS simulated with these two models reproduced the observed SFS well (Fig. **S11**), suggesting that the recovered demographic parameters were good estimates of the past population history. The estimated demographic parameters from the second best-fitting model (Model 3) overlapped better with those estimated from the Stairway Plot 2 analysis than the first best-fitting model, in terms of N_e and timing of population size

changes (Table S4). This model estimated an ancestral population size N_e of 9.5×10^3 individuals (95% CI: 9.15×10^3–1.48×10^4), which diverged into the Spanish and the ancestral population of the Chinese and Scandinavian groups at 6.20 Mya (95% CI: 5.17-6.20 Mya; Fig. **4b**, Table S4). Both populations expanded after the split. The divergence time between the Chinese and Scandinavian populations was estimated at 0.57 Mya (95% CI: 0.49-0.63 Mya; Fig. **4b**, Table S4), followed by expansion. All three populations experienced an instantaneous growth at 60.75 kya (95% CI: 27.95–88.07 kya), where N_e reached 6×10^5, 3.76×10^5, and 9.20×10^5 for the Chinese, Spanish and Scandinavian groups, respectively (Fig. **4b**, Table S4). The historical gene flow among the three groups was only 0–0.71 migrants per generation, while the current gene flow among them was as high as 30–211 migrants per generation (Fig. **4b**, Table S4).

The best-fitting model (Model 7, Table S4) uncovered a similar demographic history with estimates of N_e, divergence time, and gene flow among groups in the same orders of magnitude as those of the first model. However, the order of population split was different, as it inferred that the Chinese group split first from the ancestral population. The current N_e of the Chinese, Spanish and Scandinavian groups were estimated to be 6.60×10^5, 9.97×10^5, and 2.40×10^5, respectively (Table S4), corresponding to a Spanish N_e twice as high as in the model 3, and a Scandinavian N_e divided by three.

Contrary to expectation, we did not find any correlation between the π_0/π_4 ratio and Tajima's D, indicating no increase of genetic load in the expanding northern populations (Fig. **S12**). Additionally, neither the recessive nor the additive genetic loads calculated based on gene effects of SNP annotations are correlated with latitude or longitude (Fig. **S13**).

Genotype and environment association

Based on the GF ranking we selected the top 21 environmental variables (Table S5 & see M&M) to perform the RDA. Forward selection retained 10 of these variables (see Table S5), which were used to produce the environmental distance matrix. For the geographic distance matrix, the forward selection retained 21 of the 64 PCNM-axes. The variance partitioning showed that environment, geographic distance and coancestry together explained 19.6% of the total genetic variance, of which 10.9% was confounded among them (Fig. **5a**). The environmental matrix alone could only explain 0.88% of the total genetic variance. Only the first RDA-axis was identified as significant, and 10 SNPs appeared to be outside of the general clustering and had very low *p*-values (Fig. **5b,c**). However, none deviated from the expected distribution of *p*-values (Fig. **5c**), meaning that these SNPs are probably false

positives. The seven top SNPs appear to have been influenced mainly by arid index, nitrogen content and bio4 (standard deviation of the monthly mean temperatures).

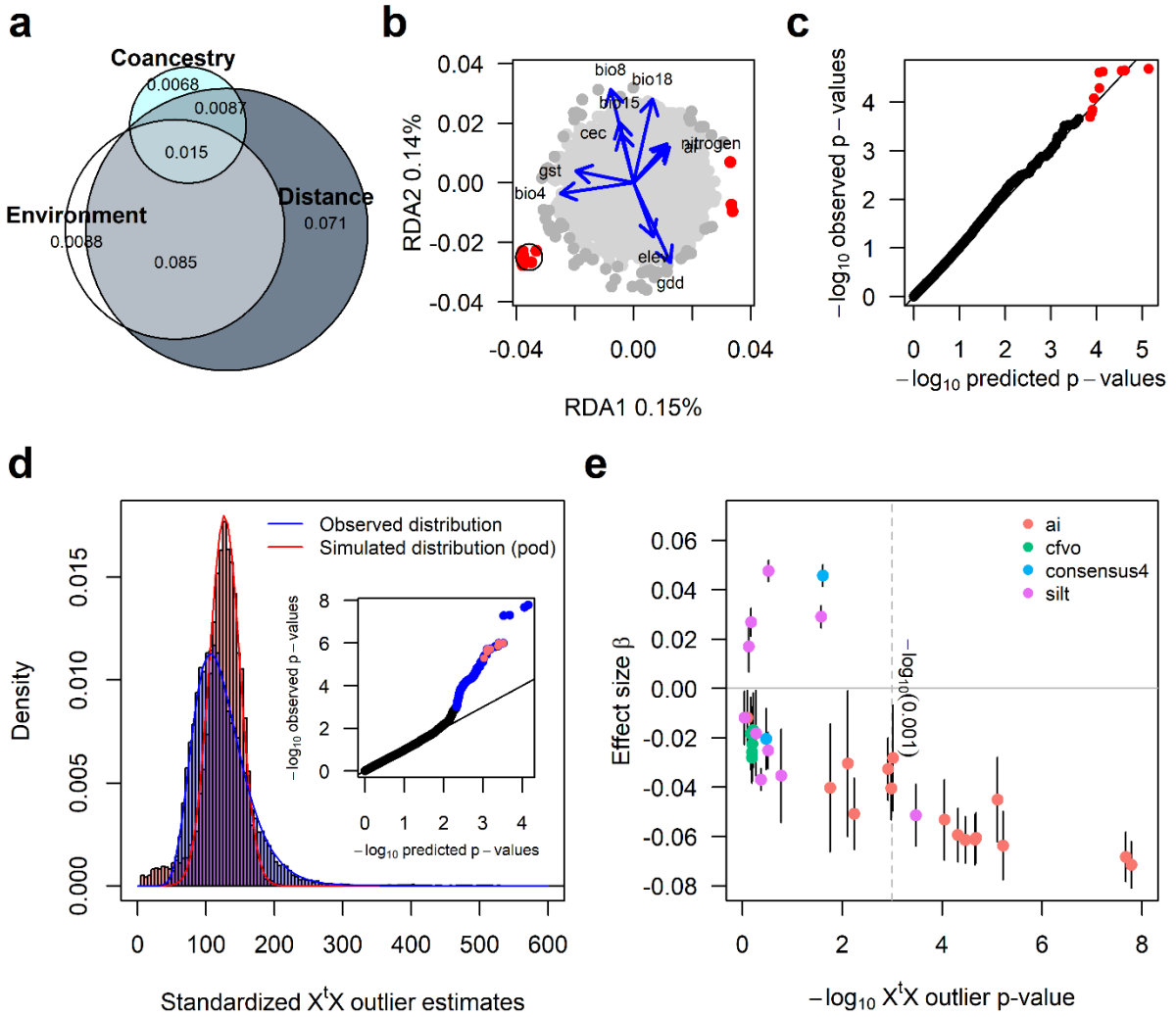


Figure 5: Genotype environment association analyses. Panel **a**: Venn diagram showing the partition of the observed total variance into geographic, co-ancestry and environmental contributions, with the three factors jointly explain 19.5%, while 7.1%, 0.88% and 0.68% independently. Panel **b**: RDA showing the top 10 SNP outliers (●), the circled 7 SNPs were found as X^tX-outliers in panel **d**. Panel **c**, no p-values of the SNPs from the RDA deviated from the expected distribution, suggesting that they are likely false positives. Panel **d**: Comparison of simulated outlier values based on the pseudo-observed data (POD) with observed values. POD estimates are based on the posterior mean and standard deviation of the two parameters of the Beta prior distribution assumed for the (across populations) frequencies of the SNP reference allele of the core model, see material and methods. Inset QQ-plot shows observed vs expected p-values given the log-normal distribution of observed X^tX outlier estimates (X^tX values > 311.93). Blue colored dots (● 103) in the inset figures are significant outlier SNPs at $p < 0.001$ and red highlighted dots correspond to those circled in panel B. Panel **e**: Loci with significant effect size of four climatic/soil variables from the multivariate BayPass analysis. Four variables were found to be associated with two or more loci. Significance was determined when the absolute value of allele effect size (β) minus its standard deviation was larger than zero, $|\beta|-\sigma(\beta) > 0$, which for this data corresponds to a posterior inclusion probability

(PIP) of 0.55 or larger. Because the variables were scaled to a unit's variance, an allele effect size of -0.06 of ai means that the arid index is 0.06 standard deviations lower when the alternative allele is present. X^tX outliers are loci that have differentiated allele frequencies across populations.

In the BayPass analysis, we first identified 103 SNPs strongly differentiated among populations (outliers), of which seven were also identified in the RDA-analysis (circled in Fig. **5b**). We then detected 33 SNPs (including 11 outliers) associated with four environmental variables: aridity index (ai), proportion of silt particles in the fine soil fraction, volumetric fraction of coarse soil fragments (cfvo), and presence of mixed forest (consensus4; Fig. **5e**, Table S6 identified by ^{B}). They all deviated from the expected distribution of *p*-values (Fig. **5d**), meaning that these 33 SNPs are probably true positives. Among these 33 SNPs, 17 were in annotated genes whose functions cover synthesis of organic compounds and oligosaccharides, intramembrane protease, splicing and processing of RNA, cell wall reconstruction and stress resistance. However, we want to emphasize that gene annotations are limited for *P. taeda* and absent for *P. sylvestris*, which strongly limits any further interpretation.

Discussion

Uniform genetic diversity across Eurasia

Scots pine was able to rapidly expand and (re)establish its distribution across the entire Eurasian continent in the post LGM period. Rapid expansion from a refugial population is expected to leave a spatial pattern characterized by declining genetic diversity and increasing genetic load due to drift and increasing inbreeding on the leading edge of the colonization front. However, our findings in Scots pine deviate from these expectations. Firstly, while there are observable patterns in the diversity (both π and H_o) along latitudes and longitudes, the magnitude of these differences is minimal. The genetic diversity is only slightly reduced in northern Europe and Asia compared to the rest of its distribution, indicating a weak founder effect originating from recolonization. Secondly, the inbreeding coefficient remains close to 0 in most populations, questioning the anticipated increase of inbreeding in geographically marginal. Thirdly, there is no discernable signal of genetic load in expanding populations.

The nucleotide diversity is overall high (mean π= 0.0062) compared to other *Pinus* and conifer species (mean π= 0.0013 – 0.0037; Eckert *et al.*, 2013; Acosta *et al.*, 2019; Shalev *et al.*, 2022; Guo *et al.*, 2023), or to other boreal tree species with large distribution like the Balsam Poplar, *Populus balsamifera* (mean π= 0.0027; Keller *et al.*, 2010), although this difference

could partly be due to different sequencing and analysis methods (Korunes & Samuk, 2021). The average observed heterozygosity (H_o= 0.258) is also high but similar to other Asian pine species (Xia *et al.*, 2018; Zhao *et al.*, 2020).

Another interesting result is the low genetic differentiation in Scots pine, with a global F_{ST} of 0.048 across its continental distribution. This is in agreement with previous results that estimated a F_{ST} of 0.02 on average for the western part of Europe (Pyhäjärvi *et al.*, 2020; Tyrmi *et al.*, 2020; Wachowiak *et al.*, 2022a; Milesi *et al.*, 2023). We confirm the genetic proximity of the Scottish populations to the mainland ones (Wachowiak *et al.*, 2011; Wachowiak *et al.*, 2022a), and the presence of genetically distinct populations in Turkey and the Carpathian Mountains detected at high *K* values (Naydenov *et al.*, 2007; Dering *et al.*, 2017; Tóth *et al.*, 2017b). The uniqueness of these populations, and in general the differentiation between European populations, are more visible at the maternally inherited mitochondrial DNA markers due to the more limited dispersal ability of seeds (Tóth *et al.*, 2017a). However, gene flow mediated by pollen is sufficiently strong to overcome seed dispersal limitations and to blur the ancestral information in the nuclear genomes of the extant populations.

Associated with the low F_{ST}, we observed a pattern of IBD, suggesting that spatial distance is a stronger barrier to gene flow than any landforms. Nonetheless, the feems analysis detected barriers to gene flow precisely where we can expect a decrease in pollen exchange: over large bodies of water, or over wide spaces of steppe or broad-leaved woods. The relative isolation of the Spanish and the Caucasian populations identified by feems confirms previous results obtained using other genetic markers (Soranzo *et al.*, 2000; Naydenov *et al.*, 2007; Semerikov *et al.*, 2020; Dering *et al.*, 2021; Wachowiak *et al.*, 2022b) and can partially be linked to counteracting dominant winds (Kling & Ackerly, 2020). These barriers, even if detectable, are overall faint, and Scots pine can be considered close to panmictic.

This is in stark contrast with Norway spruce that has a distribution largely overlapping the western part of the Scots pine but shows a strong population structure (Chen *et al.*, 2019; Milesi *et al.*, 2019). Both Norway spruce and Scots pine are wind pollinated with wind dispersed seeds and are mostly outcrossing (Burczyk *et al.*, 2004; Robledo-Arnuncio & Gil, 2005; Piotti *et al.*, 2009). One reason for their different genetic patterns could be the heavier spruce pollen compared to pine (Di-Giovanni & Kevan, 1991), which reduces the gene flow distance and enhances population genetic differentiation. Other differences in demographic history, number of LGM refugia and ecological amplitude of these species could all impact the distribution of genetic diversity in extant populations resulting in species-specific patterns of genetic structure (Chen *et al.*, 2019; Wang *et al.*, 2020; Milesi *et al.*, 2023).

Low demographic impact of the LGM

One unexpected result from our demographic analyses was the relatively stable effective population size since 60 kya and little fluctuation during the LGM. The stairwayplot2 analysis and the model 3 of the fastsimcoal2 analysis produced similar results regarding the timing of major events and final population sizes. However, the best scenario for the fastsimcoal2 analysis (model 7) inferred a split of the Chinese population before the separation of the Scandinavian and Spanish groups. Although this scenario is plausible, considering the presence of the Scots pine sister species, *P. densiflora,* in the eastern margin of Scots pine range (Saladin *et al.*, 2017; Zeb *et al.*, 2020; Jin *et al.*, 2021), the inferred effective population sizes, with a Spanish N_e four times larger than the Scandinavian N_e, seem implausible given the lower connectivity of the Spanish populations to the rest of the distribution visible in both the PCA and feems results. We therefore consider that model 3, even if ranked second, is the most plausible based on biological data.

The limited impact of the LGM on the species demography may result from either a glacial period too short in duration compared to the evolutionary timescale of the species to exert a significant influence, or from the existence of multiple glacial refugia which could have conserved the entire species genetic variation of the species. The demography of Scots pine recovered in this study supports strong changes in population size dating back to much older period in Pleistocene or Pliocene (Pyhäjärvi *et al.*, 2007; Kujala & Savolainen, 2012; Milesi *et al.*, 2023). Similar ancient changes in population size are reported for at least seven other wind pollinated tree species, including three conifers (*Picea abies*, *Pinus pinaster*, *Pinus tabuliformis*) and four angiosperms (*Betula pendula*, *Fagus sylvatica*, *Populus nigra* and *Quercus petraea*) (Xia *et al.*, 2018; Milesi *et al.*, 2023).

The demographic resilience of Scots pine over the last ice age is not entirely surprising for a cold-resistant species (Hurme *et al.*, 1997; Savolainen *et al.*, 2004; Hall *et al.*, 2021). It faces strong competition from deciduous species when the climate becomes milder (Galiano *et al.*, 2010; Boisvert-Marsh & Blois, 2021), making it probably more sensitive to interglacial than to glacial periods. Patches of trees were able to grow on surface debris accumulating on ice (Fickert *et al.*, 2007; Zale *et al.*, 2018), suggesting that the presence of trees was not limited to ice-free areas during the LGM. If Scots pine had persisted on nunataks or on ice, this would explain the absence of a strong founder effect in our data, as would be expected for a recolonizing species, the mild impact of the LGM detected on the species demography, and the very low differentiation between populations in general. This would also explain the pollen and macrofossil evidence of Scots pines presence at high latitudes very rapidly after and even during

the LGM (Huntley & Birks, 1983; Kullman, 2002; Parducci *et al.*, 2012; Zale *et al.*, 2018). The old image of a few restricted refugia during the LGM is likely inaccurate, and the Quaternary could have been much more forested than initially thought.

Contrary to the expectation of lower efficacy of purging deleterious mutations in rapidly expending populations, our analyses suggest that the fitness of Scots pine is probably not reduced at its expansion limit. Extensive gene flow plays a significant role in shaping these genetic patterns, mitigating the impact of drift and maintaining a large effective population size. Consequently, the success of Scots pine in colonizing new areas after LGM likely arises from a combination of factors, including its resilience, long-distance dispersal ability, and adaptability.

Signals of GEA

The low genetic differentiation in Scots pine is in stark contrast to the marked phenotypic differences observed among populations, including traits display very high Q_{ST} (equivalent of F_{ST} for phenotypic traits) such as bud set (Q_{ST} 0.860; Savolainen *et al.*, 2004), growth rate (Q_{ST} 0.710; Notivol *et al.*, 2007), and cold hardiness (Q_{ST} 0.82; Hall *et al.*, 2021), demonstrating strong local adaptation (Carlisle & Brown, 1968; Rehfeldt *et al.*, 2002). However, our study revealed only weak signals of genetic adaptation to environmental gradients. The overall pattern of isolation by environment is weak, as RDA identified that barely 1% of the genetic differentiation among populations could be explained by environment alone, and only 33 out of 25.8K SNPs strongly associated with environmental variables. The missing signals of GEA could result from several factors. First, local adaptation is a complex trait controlled by multiple factors that collectively provide a fitness advantage for populations in specific environments. Identifying the genetic basis of such complex traits in wild populations requires dense sampling of the genome and unrealistic sample sizes (Gienapp, 2020). In this study, we surveyed only around 0.01% of the genome, thus the possibility of detecting outliers of significant effects is slim. Similarly, few environment associated loci are detected in Scots pine in a study that analyzed 60,000 open reading frames (Tyrmi *et al.*, 2020). Second, traits under selection are likely driven by subtle changes in allele frequencies across many loci, making their detection difficult with the current methods (Nadeau *et al.*, 2016; Barton *et al.*, 2019; Hall *et al.*, 2021). Third, more intricate processes like methylation, gene expression regulation, or retrotransposon activation, which are not captured by SNP-based analyses, could play a role in local adaptation (Voronova *et al.*, 2014; Alakärppä *et al.*, 2018). Fourth, even though the LGM did not impact the species demography much, older events (e.g. around 60 kya) could overshadow signal of

selection and GEA. This complexity is further exacerbated when colonization routes covary with climate gradients along latitudes or longitudes (Zhao *et al.*, 2020; Capblancq *et al.*, 2023).

Nonetheless, among the four significant environmental variables that drove the few allele frequency changes in this study, the most important one was the aridity index. This is consistent with the ecology of Scots pine, which is known to suffer from drought, resulting in a reduction in tree height in the southern part of the distribution (Hallingbäck *et al.*, 2021), shorter and fewer shoots (Taeger *et al.*, 2013; Bachofen *et al.*, 2021), and increased mortality following severe drought events (Haberstroh *et al.*, 2022). Our results highlight the challenges of detecting selection wild populations of conifer trees. Obtaining a more comprehensive understanding of how natural selection shapes the genetic variation underlying adaptive traits will likely necessitate an approach that considers not only the genetic and environmental aspects but also refines the phenotypic contexts of adaptation under reduced dimensionality of environmental conditions.

Acknowledgments

We are grateful to the collectors who sent us samples from the field: Philippe Castellar, Sébastien Cecchini (Office National des Forêts), Jean-Paul Collin, Véronique Guérin-Faublée, Caroline Guignier (Monteco), Cédric Hache (Lycée agricole du Haut-Languedoc), Jean-Philippe Hell (Office National des Forêts), Thierry Helminger (Musée national d'histoire naturelle de Luxembourg), Bertrand Jarri (Mayenne Nature Environnement), Isabelle Kerzaon (Université Claude Bernard Lyon 1), Laurent Lathuillière (Office National des Forêts), Marc Oesterle, Benoît Renaux (Conservatoire botanique national du Massif Central), Claude Roulet, Florent Sabatier, Hugues Savay-Guerraz, Hubert Schmuck (Office National des Forêts), Gilles Thébaud (Université Clermont-Auvergne), Hervé Tournier (Parc Naturel Régional du Vercors), Mircea Verghelet and Ionut Bordea (Piatra Craiului National Park), Natalia Demidova (NRIF), Seppo Ruotsalainen (LUKE), Arne Steffenrem and Øyvind Meland Edvardsen (NIBIO). The SAGE data were used with permission of The Center for Sustainability and the Global Environment, Nelson Institute for Environmental Studies, University of Wisconsin-Madison. The computation was enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) through High-Performance Computing Center North (HPC2N), partially funded by the Swedish Research Council through grant agreement no. 2018-05973. This study

was supported by Formas (2018-00842, 2021-02155), Carl Tryggers Stiftelse and T4F program Sweden.

Competing interests

The authors declare no competing interests.

Author contributions

XRW designed the study and participated in genotyping; JB, WZ and DH conducted genotyping and data analyses; ALC, PA, ADD, OG and HK collected samples from natural populations from their respective regions; AS and VS provided DNA samples; JB, WZ, DH and XRW wrote the manuscript draft, with input from all authors.

Data availability

The raw read data have been deposited in the NCBI database under the accession number PRJNA976641 (https://www.ncbi.nlm.nih.gov/bioproject/). The correspondence between the file names and the sampling location is available as a supplementary file.

References

Acosta JJ, Fahrenkrog AM, Neves LG, Resende MFR, Dervinis C, Davis JM, Holliday JA, Kirst M. 2019. Exome resequencing reveals evolutionary history, genomic diversity, and targets of selection in the conifers *Pinus taeda* and *Pinus elliottii*. *Genome Biology and Evolution* **11**(2): 508-520.

Akaike H. 1987. Factor analysis and AIC. *Psychometrika* **52**(3): 317–332.

Alakärppä E, Salo HM, Valledor L, Cañal MJ, Häggman H, Vuosku J. 2018. Natural variation of DNA methylation and gene expression may determine local adaptations of Scots pine populations. *Journal of experimental botany* **69**(21): 5293-5305.

Andersson B, Fedorkov A. 2004. Longitudinal differences in Scots pine frost hardiness. *Silvae Genetica* **53**: 76-80.

Bachofen C, Perret-Gentil A, Wohlgemuth T, Vollenweider P, Moser B. 2021. Phenotypic plasticity versus ecotypic differentiation under recurrent summer drought in two drought-tolerant pine species. *Journal of Ecology* **109**(11): 3861-3876.

Barton N, Hermisson J, Nordborg M. 2019. Population genetics: Why structure matters. *eLife* **8**: e45380.

Boisvert-Marsh L, Blois Sd. 2021. Unravelling potential northward migration pathways for tree species under climate change. *Journal of Biogeography* **48**(5): 1088-1100.

Borcard D, Legendre P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological modelling* **153**(1-2): 51-68.

Burczyk J, DiFazio SP, Adams WT. 2004. Gene flow in forest trees: how far do genes really travel? *Forest Genetics* **11**(3/4): 179.

Capblancq T, Lachmuth S, Fitzpatrick MC, Keller SR. 2023. From common gardens to candidate genes: exploring local adaptation to climate in red spruce. *New Phytologist* **237**(5): 1590-1605.

Carlisle A, Brown A. 1968. *Pinus sylvestris* L. *Journal of Ecology* **56**(1): 269-307.

Caye K, Deist TM, Martins H, Michel O, François O. 2016. TESS3: fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources* **16**(2): 540-548.

Cheddadi R, Vendramin GG, Litt T, François L, Kageyama M, Lorentz S, Laurent J-M, Beaulieu J-LD, Sadori L, Jost A, et al. 2006. Imprints of glacial refugia in the modern genetic diversity of *Pinus sylvestris*. *Global Ecology and Biogeography* **15**(3): 271-282.

Chen J, Li L, Milesi P, Jansson G, Berlin M, Karlsson B, Aleksic J, Vendramin GG, Lascoux M. 2019. Genomic data provide new insights on the demographic history and the extent of recent material transfers in Norway spruce. *Evolutionary Applications* **12**(8): 1539-1551.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *fly* **6**(2): 80-92.

Corl A, Bi K, Luke C, Challa AS, Stern AJ, Sinervo B, Nielsen R. 2018. The genetic basis of adaptation following plastic changes in coloration in a novel environment. *Current Biology* **28**(18): 2970-2977. e2977.

Davis MB, Shaw RG. 2001. Range shifts and adaptive responses to Quaternary climate change. *Science* **292**(5517): 673-679.

de Pedro M, Riba M, González-Martínez SC, Seoane P, Bautista R, Claros MG, Mayol M. 2021. Demography, genetic diversity and expansion load in the colonizing species Leontodon longirostris (Asteraceae) throughout its native range. *Molecular Ecology* **30**(5): 1190-1205.

Dering M, Baranowska M, Beridze B, Chybicki IJ, Danelia I, Iszkuło G, Kvartskhava G, Kosiński P, Rączka G, Thomas PA, et al. 2021. The evolutionary heritage and ecological uniqueness of Scots pine in the Caucasus ecoregion is at risk of climate changes. *Scientific Reports* **11**(1): 22845.

Dering M, Kosiński P, Wyka TP, Pers–Kamczyc E, Boratyński A, Boratyńska K, Reich PB, Romo A, Zadworny M, Żytkowiak R, et al. 2017. Tertiary remnants and Holocene colonizers: Genetic structure and phylogeography of Scots pine reveal higher genetic diversity in young boreal than in relict Mediterranean populations and a dual colonization of Fennoscandia. *Diversity and Distributions* **23**(5): 540-555.

Di-Giovanni F, Kevan PG. 1991. Factors affecting pollen dynamics and its importance to pollen contamination: a review. *Canadian Journal of Forest Research* **21**(8): 1155-1170.

Eckert AJ, Bower AD, Jermstad KD, Wegrzyn JL, Knaus BJ, Syring JV, Neale DB. 2013. Multilocus analyses reveal little evidence for lineage-wide adaptive evolution within major clades of soft pines (*Pinus* subgenus *Strobus*). *Molecular Ecology* **22**(22): 5635-5650.

Ellegren H, Galtier N. 2016. Determinants of genetic diversity. *Nature Reviews Genetics* **17**(7): 422-433.

Ellis N, Smith SJ, Pitcher CR. 2012. Gradient forests: calculating importance gradients on physical predictors. *Ecology* **93**(1): 156-168.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14**(8): 2611-2620.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLOS Genetics* **9**(10): e1003905.

Excoffier L, Foll M, Petit RJ. 2009. Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics* **40**: 481-501.

Excoffier L, Marchi N, Marques DA, Matthey-Doret R, Gouy A, Sousa VC. 2021. fastsimcoal2: demographic inference under complex evolutionary scenarios. *Bioinformatics* **37**(24): 4882–4885.

Fickert T, Friend D, Grüninger F, Molnia B, Richter M. 2007. Did debris-covered glaciers serve as Pleistocene refugia for plants? A new hypothesis derived from observations of recent plant growth on glacier surfaces. *Arctic, Antarctic, and Alpine Research* **39**(2): 245-257.

Fox EA, Wright AE, Fumagalli M, Vieira FG. 2019. ngsLD: evaluating linkage disequilibrium using genotype likelihoods. *Bioinformatics* **35**(19): 3855-3856.

Fuchs J, Jovtchev G, Schubert I. 2008. The chromosomal distribution of histone methylation marks in gymnosperms differs from that of angiosperms. *Chromosome Research* **16**: 891-898.

Galiano L, Martínez-Vilalta J, Lloret F. 2010. Drought-induced multifactor decline of Scots pine in the Pyrenees and potential vegetation change by the expansion of co-occurring oak species. *Ecosystems* **13**: 978-991.

Gamba D, Muchhala N. 2020. Global patterns of population genetic differentiation in seed plants. *Molecular Ecology* **29**(18): 3413–3428.

Gautier M. 2015. Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics* **201**(4): 1555-1579.

Gienapp P. 2020. Opinion: Is gene mapping in wild populations useful for understanding and predicting adaptation to global change? *Global Change Biology* **26**(5): 2737-2749.

Guo J-F, Wang B, Liu Z-L, Mao J-F, Wang X-R, Zhao W. 2023. Low genetic diversity and population connectivity fuel vulnerability to climate change for the Tertiary relict pine *Pinus bungeana*. *Journal of Systematics and Evolution* **61**(1): 143-156.

Haberstroh S, Werner C, Grün M, Kreuzwieser J, Seifert T, Schindler D, Christen A. 2022. Central European 2018 hot drought shifts scots pine forest to its tipping point. *Plant Biology* **24**(7): 1186-1197.

Hall D, Olsson J, Zhao W, Kroon J, Wennström U, Wang X-R. 2021. Divergent patterns between phenotypic and genetic variation in Scots pine. *Plant Communications* **2**: 100139.

Hallingbäck HR, Burton V, Vizcaíno-Palomar N, Trotter F, Liziniewicz M, Marchi M, Berlin M, Ray D, Benito Garzón M. 2021. Managing uncertainty in Scots pine range-wide adaptation under climate change. *Frontiers in Ecology and Evolution* **9**: 807.

Hamrick JL, Godt MJW. 1996. Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **351**(1345): 1291-1298.

Hewitt GM. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**(3): 247-276.

Hewitt GM. 2000. The genetic legacy of the Quaternary ice ages. *Nature* **405**(6789): 907-913.

Houston Durrant T, De Rigo D, Caudullo G. 2016. *Pinus sylvestris* in Europe: distribution, habitat, usage and threats. In: San-Miguel-Ayanz J, de Rigo D, Caudullo G, Houston Durrant T, Mauri A eds. *European Atlas of Forest Tree Species*. Luxembourg: Publications Office of the European Union, 132-133.

Huntley B, Birks HJB. 1983. *Atlas of past and present pollen maps for Europe, 0-13,000 years ago*: Cambridge University Press.

Hurme P, Repo T, Savolainen O, Pääkkönen T. 1997. Climatic adaptation of bud set and frost hardiness in Scots pine (*Pinus sylvestris*). *Canadian Journal of Forest Research* **27**(5): 716-723.

Jin W-T, Gernandt DS, Wehenkel C, Xia X-M, Wei X-X, Wang X-Q. 2021. Phylogenomic and ecological analyses reveal the spatiotemporal evolution of global pines. *Proceedings of the National Academy of Sciences* **118**(20).

Keller SR, Olson MS, Silim S, Schroeder W, Tiffin P. 2010. Genomic diversity, population structure, and migration following rapid range expansion in the Balsam Poplar, Populus balsamifera. *Molecular Ecology* **19**(6): 1212-1226.

Kling MM, Ackerly DD. 2020. Global wind patterns and the vulnerability of wind-dispersed species to climate change. *Nature Climate Change* **10**(9): 868-875.

Koenigk T, Key J, Vihma T. 2020. Climate change in the Arctic. *Physics and chemistry of the Arctic atmosphere*: 673-705.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**: 356.

Korunes KL, Samuk K. 2021. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Molecular Ecology Resources* **21**(4): 1359-1368.

Kujala ST, Savolainen O. 2012. Sequence variation patterns along a latitudinal cline in Scots pine (*Pinus sylvestris*): signs of clinal adaptation? *Tree Genetics & Genomes* **8**(6): 1451-1467.

Kullman L. 2002. Boreal tree taxa in the central Scandes during the Late-Glacial: implications for Late-Quaternary forest history. *Journal of Biogeography* **29**(9): 1117-1124.

Legendre P, Legendre L. 2012. *Numerical ecology*: Elsevier.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* doi: 10.48550/arXiv.1303.3997

Liu X, Fu Y-X. 2015. Exploring population size changes using SNP frequency spectra. *Nature Genetics* **47**(5): 555-559.

Liu X, Fu Y-X. 2020. Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biology* **21**(1): 280.

Marcus J, Ha W, Barber RF, Novembre J. 2021. Fast and flexible estimation of effective migration surfaces. *eLife* **10**: e61927.

McInnes L, Healy J, Melville J. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* doi: https://doi.org/10.48550/arXiv.1802.03426

McLean EH, Prober SM, Stock WD, Steane DA, Potts BM, Vaillancourt RE, Byrne M. 2014. Plasticity of functional traits varies clinally along a rainfall gradient in Eucalyptus tricarpa. *Plant, cell & environment* **37**(6): 1440-1451.

Meisner J, Albrechtsen A. 2018. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* **210**(2): 719-731.

Milesi P, Berlin M, Chen J, Orsucci M, Li L, Jansson G, Karlsson B, Lascoux M. 2019. Assessing the potential for assisted gene flow using past introduction of Norway spruce in southern Sweden: Local adaptation and genetic basis of quantitative traits in trees. *Evolutionary Applications* **12**(10): 1946-1959.

Milesi P, Kastally C, Dauphin B, Cervantes S, Bagnoli F, Budde KB, Cavers S, Ojeda DI, Fady B, Faivre-Rampant P, et al. 2023. Synchronous effective population size changes and genetic stability of forest trees through glacial cycles. *bioRxiv* doi: 10.1101/2023.01.05.522822

Nadeau S, Meirmans PG, Aitken SN, Ritland K, Isabel N. 2016. The challenge of separating signatures of local adaptation from those of isolation by distance and colonization history: The case of two white pines. *Ecology and Evolution* **6**(24): 8649-8664.

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 Evolutionary Biology* **7**(1): 233.

Nazareno AG, Bemmels JB, Dick CW, Lohmann LG. 2017. Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. *Molecular Ecology Resources* **17**(6): 1136-1147.

Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D, Crepeau MW, Cardeno C, Koriabine M, Holtz-Morris AE, Liechty JD, et al. 2014. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology* **15**(3): R59.

Notivol E, Garcia-Gil M, Alia R, Savolainen O. 2007. Genetic variation of growth rhythm traits in the limits of a latitudinal cline in Scots pine. *Canadian Journal of Forest Research* **37**(3): 540-551.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, et al. 2019. *vegan: community ecology package*.URL: https://CRAN.R-project.org/package=vegan

Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L. 2013. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology* **22**(24): 5983-5999.

Pan J, Wang B, Pei Z-Y, Zhao W, Gao J, Mao J-F, Wang X-R. 2015. Optimization of the genotyping-by-sequencing strategy for population genomic analysis in conifers. *Molecular Ecology Resources* **15**(4): 711-722.

Parducci L, Jørgensen T, Tollefsrud MM, Elverland E, Alm T, Fontana SL, Bennett KD, Haile J, Matetovici I, Suyama Y, et al. 2012. Glacial survival of boreal trees in northern Scandinavia. *Science* **335**(6072): 1083-1086.

Petit RJ, Aguinagalde I, de Beaulieu J-L, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, et al. 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**(5625): 1563-1565.

Piotti A, Leonardi S, Piovani P, Scalfi M, Menozzi P. 2009. Spruce colonization at treeline: where do those seeds come from? *Heredity* **103**(2): 136-145.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**(2): 945-959.

Pyhäjärvi T, García-Gil MR, Knürr T, Mikkonen M, Wachowiak W, Savolainen O. 2007. Demographic history has influenced nucleotide diversity in European *Pinus sylvestris* populations. *Genetics* **177**(3): 1713-1724.

Pyhäjärvi T, Kujala ST, Savolainen O. 2020. 275 years of forestry meets genomics in *Pinus sylvestris*. *Evolutionary Applications* **13**(1): 11-30.

Pyhäjärvi T, Salmela MJ, Savolainen O. 2008. Colonization routes of *Pinus sylvestris* inferred from distribution of mitochondrial DNA variation. *Tree Genetics & Genomes* **4**(2): 247-254.

Rehfeldt GE, Tchebakova NM, Parfenova YI, Wykoff WR, Kuzmina NA, Milyutin LI. 2002. Intraspecific responses to climate in *Pinus sylvestris*. *Global Change Biology* **8**(9): 912-929.

Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R. 2015. A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology* **24**(17): 4348-4370.

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**(1): 13-22.

Saladin B, Leslie AB, Wüest RO, Litsios G, Conti E, Salamin N, Zimmermann NE. 2017. Fossils matter: improved estimates of divergence times in *Pinus* reveal older diversification. *BMC Evolutionary Biology* **17**(1): 95.

Sannikov SN, Petrova IV. 2012. Phylogenogeography and genotaxonomy of *Pinus sylvestris* L. populations. *Russian Journal of Ecology* **43**(4): 273-280.

Savolainen O, Bokma F, García-Gil R, Komulainen P, Repo T. 2004. Genetic variation in cessation of growth and frost hardiness and consequences for adaptation of *Pinus sylvestris* to climatic changes. *Forest Ecology and Management* **197**(1-3): 79-89.

Savolainen O, Lascoux M, Merila J. 2013. Ecological genomics of local adaptation. *Nature Reviews Genetics* **14**(11): 807-820.

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

Semerikov NV, Petrova IV, Sannikov SN, Semerikova SA, Tashev AN, Lascoux M, Semerikov VL. 2020. Cytoplasmic DNA variation does not support a recent contribution of *Pinus sylvestris* L. from the Caucasus to the main range. *Tree Genetics & Genomes* **16**(4): 59.

Shalev TJ, El-Dien OG, Yuen MM, Shengqiang S, Jackman SD, Warren RL, Coombe L, van der Merwe L, Stewart A, Boston LB. 2022. The western redcedar genome reveals low genetic diversity in a self-compatible conifer. *Genome Research* **32**(10): 1952-1964.

Skotte L, Korneliussen TS, Albrechtsen A. 2013. Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**(3): 693-702.

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. *Molecular Ecology* **9**(9): 1205-1211.

Taeger S, Fussi B, Konnert M, Menzel A. 2013. Large-scale genetic structure and drought-induced effects on European Scots pine (*Pinus sylvestris* L.) seedlings. *European Journal of Forest Research* **132**(3): 481-496.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**(3): 585-595.

Tóth EG, Bede-Fazekas Á, Vendramin GG, Bagnoli F, Höhn M. 2019. Mid-Pleistocene and Holocene demographic fluctuation of Scots pine (*Pinus sylvestris* L.) in the Carpathian Mountains and the Pannonian Basin: Signs of historical expansions and contractions. *Quaternary International* **504**: 202-213.

Tóth EG, Köbölkuti ZA, Pedryc A, Höhn M. 2017a. Evolutionary history and phylogeography of Scots pine (*Pinus sylvestris* L.) in Europe based on molecular markers. *Journal of Forestry Research* **28**(4): 637-651.

Tóth EG, Vendramin GG, Bagnoli F, Cseke K, Höhn M. 2017b. High genetic diversity and distinct origin of recently fragmented Scots pine (*Pinus sylvestris* L.) populations along the Carpathians and the Pannonian Basin. *Tree Genetics & Genomes* **13**(2): 47.

Tyrmi JS, Vuosku J, Acosta JJ, Li Z, Sterck L, Cervera MT, Savolainen O, Pyhäjärvi T. 2020. Genomics of clinal local adaptation in *Pinus sylvestris* under continuous environmental and spatial genetic setting. *G3: Genes, Genomes, Genetics* **10**(8): 2683-2696.

Vitasse Y, Bresson CC, Kremer A, Michalet R, Delzon S. 2010. Quantifying phenological plasticity to temperature in two temperate tree species. *Functional ecology* **24**(6): 1211-1218.

Voronova A, Belevich V, Jansons A, Rungis D. 2014. Stress-induced transcriptional activation of retrotransposon-like sequences in the Scots pine (*Pinus sylvestris* L.) genome. *Tree Genetics & Genomes* **10**(4): 937-951.

Wachowiak W, Perry A, Zaborowska J, Gonzalez-Martinez SC, Cavers S. 2022a. Admixture and selection patterns across the European distribution of Scots pine, *Pinus sylvestris* (Pinaceae). *Botanical Journal of the Linnean Society* **200**(3): 416–432.

Wachowiak W, Salmela MJ, Ennos RA, Iason G, Cavers S. 2011. High genetic diversity at the extreme range edge: nucleotide variation at nuclear loci in Scots pine (*Pinus sylvestris* L.) in Scotland. *Heredity* **106**(5): 775-787.

Wachowiak W, Żukowska WB, Perry A, Lewandowski A, Cavers S, Łabiszak B. 2022b. Phylogeography of Scots pine in Europe and Asia based on mtDNA polymorphisms. *Journal of Systematics and Evolution* **61**(2): 315-327.

Wang IJ, Bradburd GS. 2014. Isolation by environment. *Molecular Ecology* **23**(23): 5649-5662.

Wang X, Bernhardsson C, Ingvarsson PK. 2020. Demography and natural selection have shaped genetic variation in the widely distributed conifer Norway Spruce (*Picea abies*). *Genome Biology and Evolution* **12**(2): 3803-3817.

Willis KJ, Rudner E, Sümegi P. 2000. The full-glacial forests of central and southeastern Europe. *Quaternary Research* **53**(2): 203-213.

Willis KJ, van Andel TH. 2004. Trees or no trees? The environments of central and eastern Europe during the Last Glaciation. *Quaternary Science Reviews* **23**(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**(1): 90-101.

Wright S. 1943. Isolation by distance. *Genetics* **28**(2): 114.

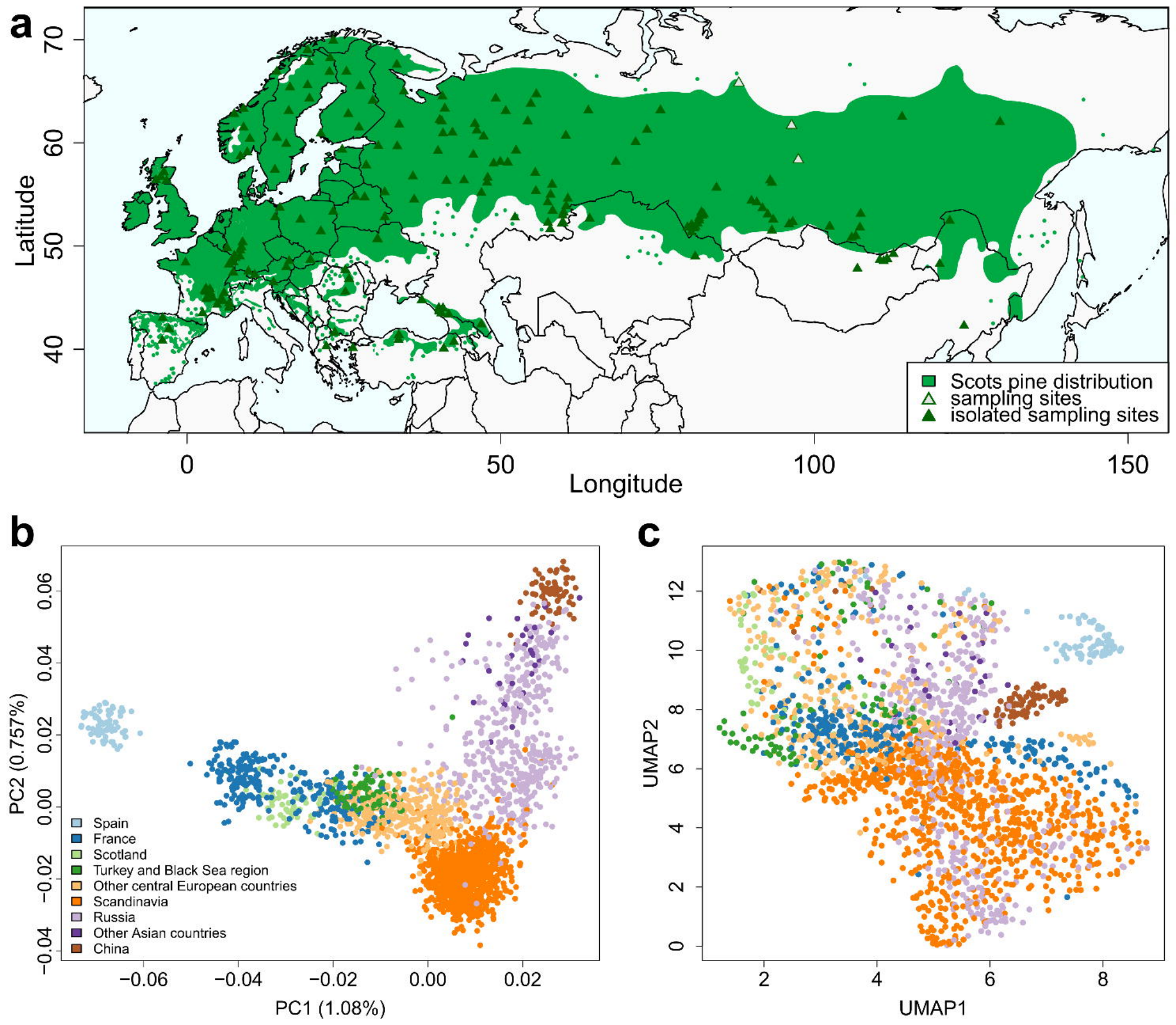
Xia H, Wang B, Zhao W, Pan J, Mao J-F, Wang X-R. 2018. Combining mitochondrial and nuclear genome analyses to dissect the effects of colonization, environment, and geography on population structure in *Pinus tabuliformis*. *Evolutionary Applications* **11**(10): 1931-1945.

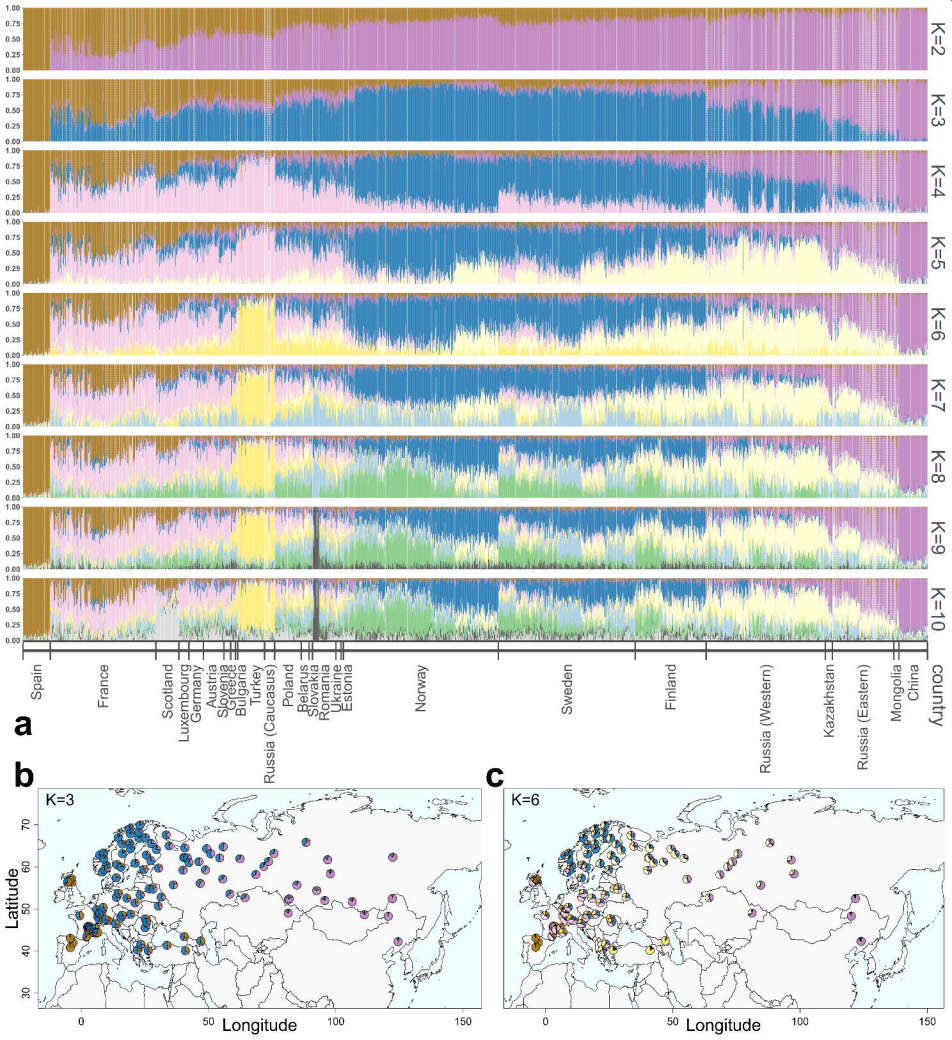
Zale R, Huang Y-T, Bigler C, Wood JR, Dalén L, Wang X-R, Segerström U, Klaminder J. 2018. Growth of plants on the Late Weichselian ice-sheet during Greenland interstadial-1? *Quaternary Science Reviews* **185**: 222-229.

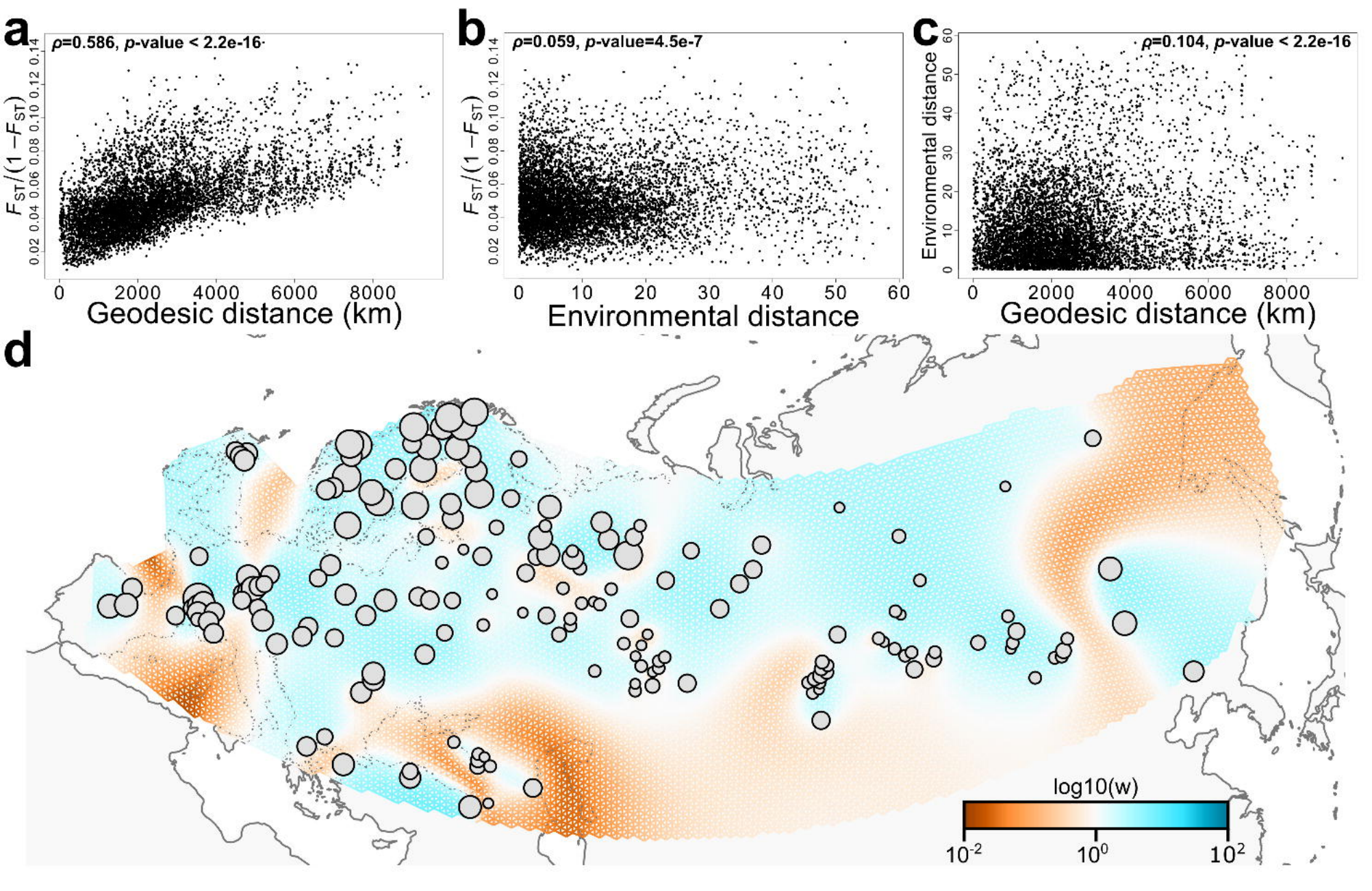
Zeb U, Dong W-L, Zhang T-T, Wang R-N, Shahzad K, Ma X-F, Li Z-H. 2020. Comparative plastid genomics of *Pinus* species: Insights into sequence variations and phylogenetic relationships. *Journal of Systematics and Evolution* **58**(2): 118-132.

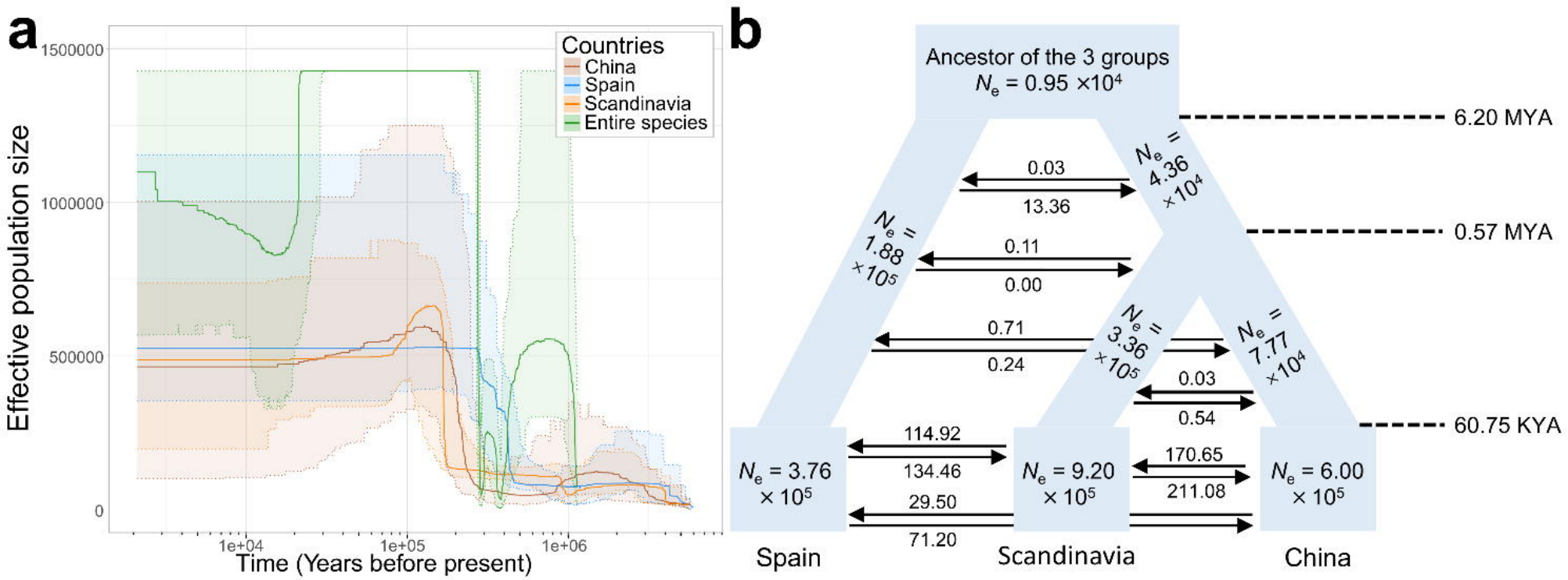
Zhao W, Sun Y-Q, Pan J, Sullivan AR, Arnold ML, Mao J-F, Wang X-R. 2020. Effects of landscapes and range expansion on population structure and local adaptation. *New Phytologist* **228**(1): 330-343.

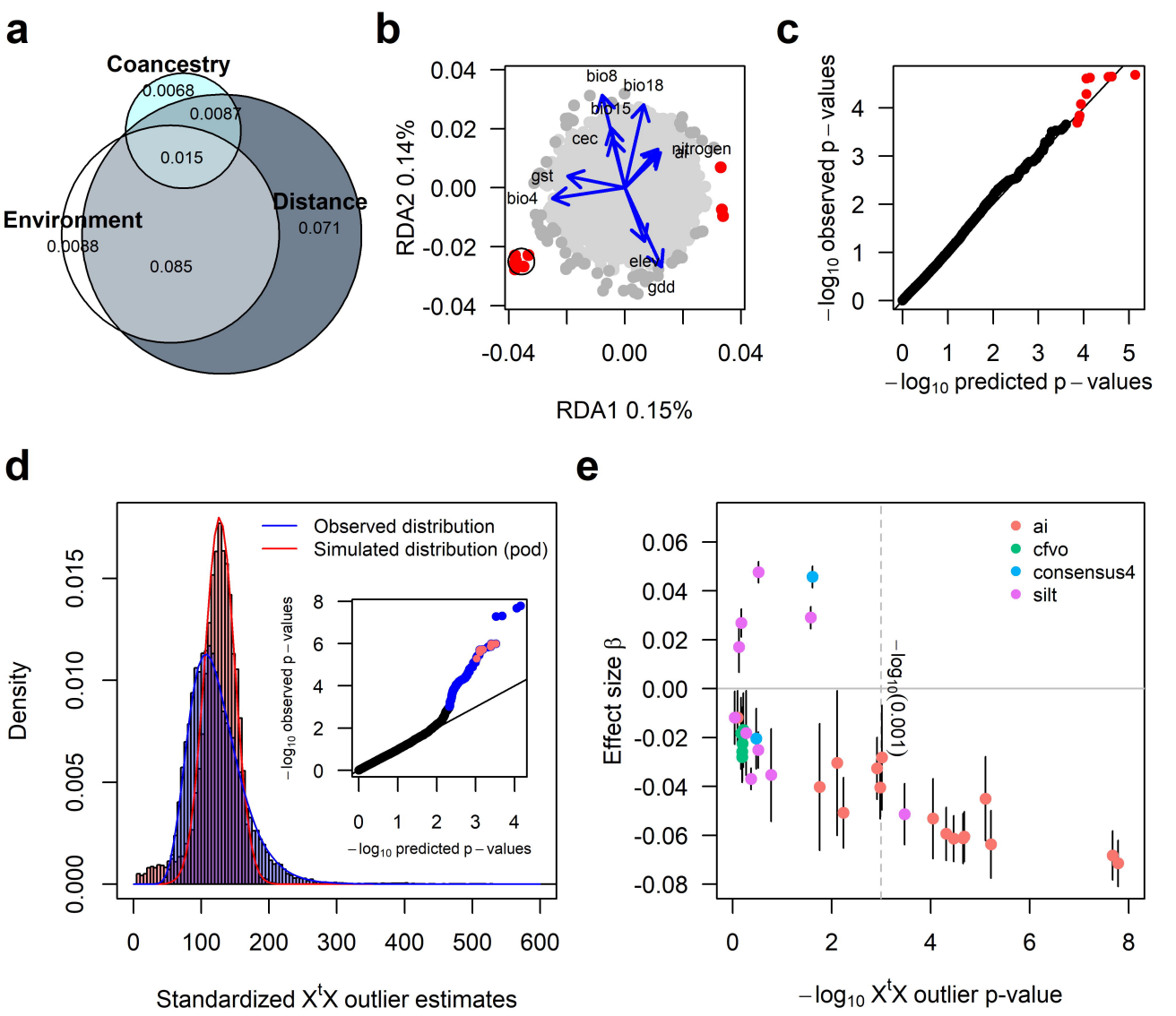
Zimin A, Stevens KA, Crepeau MW, Holtz-Morris A, Koriabine M, Marçais G, Puiu D, Roberts M, Wegrzyn JL, Jong PJd, et al. 2014. Sequencing and assembly of the 22-Gb Loblolly pine genome. *Genetics* **196**(3): 875-890.


a
Latitude
70
60
50
40
0
50
100
150
Longitude
Scots pine distribution
sampling sites
isolated sampling sites
b
PC2 (0.757%)
0.06
0.04
0.02
0.00
−0.02
−0.04
−0.06
−0.04
−0.02
0.00
0.02
PC1 (1.08%)
Spain
France
Scotland
Turkey and Black Sea region
Other central European countries
Scandinavia
Russia
Other Asian countries
China
c
UMAP2
12
10
8
6
4
2
0
2
4
6
8
UMAP1


K=2
K=3
K=4
K=5
K=6
K=7
K=8
K=9
K=10
1.00
0.75
0.50
0.25
0.00
Spain
France
Scotland
Luxembourg
Germany
Austria
Slovenia
Greece
Bulgaria
Turkey
Russia (Caucasus)
Poland
Belarus
Slovakia
Romania
Ukraine
Lithuania
Estonia
Norway
Sweden
Finland
Russia (Western)
Kazakhstan
Russia (Eastern)
Mongolia
China
country
a
b
K=3
Latitude
Longitude
c
K=6
Longitude


a
ρ=0.586, p-value < 2.2e-16
$F_{ST}/(1-F_{ST})$
0.02 0.04 0.06 0.08 0.10 0.12 0.14
0 2000 4000 6000 8000
Geodesic distance (km)
b
ρ=0.059, p-value=4.5e-7
$F_{ST}/(1-F_{ST})$
0.02 0.04 0.06 0.08 0.10 0.12 0.14
0 10 20 30 40 50 60
Environmental distance
c
ρ=0.104, p-value < 2.2e-16
Environmental distance
0 10 20 30 40 50 60
0 2000 4000 6000 8000
Geodesic distance (km)
d
log10(w)
10^{-2} 10^{0} 10^{2}


a
Countries
China
Spain
Scandinavia
Entire species
Effective population size
1500000
1000000
500000
0
1e+04
1e+05
1e+06
Time (Years before present)
b
Ancestor of the 3 groups
N_e = 0.95 ×10⁴
6.20 MYA
0.03
13.36
N_e = 4.36 ×10⁴
0.57 MYA
N_e = 1.88 ×10⁵
0.11
0.00
0.71
0.24
N_e = 3.36 ×10⁵
N_e = 7.77 ×10⁴
0.03
0.54
60.75 KYA
114.92
134.46
N_e = 3.76 × 10⁵
N_e = 9.20 × 10⁵
170.65
211.08
N_e = 6.00 × 10⁵
29.50
71.20
Spain
Scandinavia
China


a
Coancestry
0.0068
0.0087
0.015
Environment
0.0088
0.085
Distance
0.071
b
bio8
bio18
bio15
cec
nitrogen
air
gst
bio4
ele
gdd
RDA2 0.14%
RDA1 0.15%
c
−log10 observed p − values
−log10 predicted p − values
d
Observed distribution
Simulated distribution (pod)
Density
Standardized X^tX outlier estimates
e
ai
cfvo
consensus4
silt
−log10(0.001)
Effect size β
−log10 X^tX outlier p-value