Tree Genetics & Genomes (2025) 21:5
https://doi.org/10.1007/s11295-025-01689-7

ORIGINAL ARTICLE

Genomic exploration of *Thuja plicata* indicates no loss of diversity from adults to offspring in UK Woodlands

Laura Guillardín[1,3] · Ella Glover[1] · Gary Kerr[2] · John J. MacKay[1]

Received: 20 March 2024 / Revised: 30 December 2024 / Accepted: 12 January 2025 / Published online: 3 February 2025
© The Author(s) 2025

Abstract

Thuja plicata is a conifer tree that is valued for its cultural, ecological and wood quality features in its natural range in western North America and is used in Europe as an exotic timber species. It is increasingly used in mixed species forests in the UK that are managed using natural regeneration, raising questions about genetic diversity of both the adult trees and offspring of the species in these stands. We studied population structure and genetic diversity in four UK woodlands of both adults and naturally regenerated offspring of *T. plicata*. We discovered 194,154 Single Nucleotide Polymorphisms (SNPs) using Genotyping-by-Sequencing (GBS) and retained 655 SNPs across 598 genome sequence scaffolds for analysis. The majority of these scaffolds included only one SNP, suggesting that; the selected SNPs were widely distributed within the genome. We found many monomorphic sites, most of which were restricted to adults in a single woodland, indicative of a genetic differentiation among woodlands. We found three ancestral populations (K = 3) and low levels of admixture across the four sites, indicating that the starting materials were largely from single populations. Estimates of genetic diversity using heterozygosity (H_o) and nucleotide diversity (π) were low (overall $H_o = 0.174$, $\pi = 0.00226$) across all sites but slightly higher in juveniles at two sites. Our genotyping methods and results on standing genetic diversity in this important conifer provide insight into the ability of planted woodlands in the UK to face environmental shifts and disease threats.

Keywords *Thuja plicata* · Genetic diversity · Genomic exploration · Single nucleotide polymorphism · Natural regeneration *Thuja plicata*

Introduction

Genetic diversity is the basis for adaptation and is an indicator of evolutionary potential (Sgrò et al. 2011; Wernberg et al. 2018). Conservation of the genetic diversity of species has become a priority to protect and enhance the diversity of ecosystems (Hoban et al. 2020, 2023; Heuertz et al. 2023). Tree species are long-lived and highly genetically diverse due to their out-crossing nature. Forestry species used in commercial production are only at the early stages of domestication (Savolainen and Pyhäjärvi 2007). However, climate change may surpass the potential for adaptation in forestry species (Alberto et al. 2013), which could affect populations in both the natural range and in areas of artificial introduction.

In forestry species, it is a widespread practice to select seed sources among populations from across the native distribution to establish plantations or new populations. Several studies have validated the use of specific provenances and populations by developing trials evaluating the population growth performance under different environmental conditions (Pierro et al. 2017; Baldi and Porta 2022). This may lead to over-sampling specific provenances to establish plantations outside of the natural range, and thus narrowing the gene pool in the recipient woodlands. Therefore, there is a risk that genetic diversity of non-native plantations may be

Communicated by C. Rellstab

✉ Laura Guillardín
laura.guillardin@teagasc.ie

[1] Department of Biology, University of Oxford, Oxford OX1 3RB, UK

[2] Forest Research, Alice Holt Lodge, Farnham, Surrey GU10 4LH, UK

[3] Present address: Forestry Development Department, Teagasc, Athenry, Co., Galway H65R718, Ireland

lower compared to the genetic diversity of the same species in their natural distribution. Public and private woodland estates may hold records of the material originally used for the establishment of their plantations. Although, this information may be lacking, especially considering initial establishment may be several decades to centuries in the past, creating uncertainty concerning the suitability and long-term viability of non-native plantations.

Woodland cover in the UK is only 14%, with approximately 25% of this area accounted for by a single species (*Picea sitchensis*), mostly comprised of single species plantations (Forest research 2022). *Thuja plicata* Donn ex D.Don is found in < 1% of the forested area, however this is predicted to increase due to the desire to diversify tree species used in forestry as an adaptation to climate change (Gil-Moreno et al. 2016). *Thuja plicata* is a large evergreen tree from the Cupressaceae family, native to western North America where it has been used and managed by First Nations and Native Americans for centuries for cultural and utility items (Hebda and Mathewes 1984). Due to its valuable wood characteristics such as durability and resistance to rot (Stirling et al. 2017), it is recognised as producing high quality timber for construction and is widely commercially planted for this very reason.

Thuja plicata is a shade tolerant species, which increases its potential use in mixed species forest, however, recommendations on the use of specific provenances and information on the origin and genetic makeup of UK plantations are either lacking or scarce (Kerr 2019). A study of various growth traits in the UK among provenances across the coastal region showed higher growth performance in materials originating between 46°N and 50°N of Washington and British Columbia (Kerr 2019). *Thuja plicata* is monoecious, i.e. individuals have both the male and female reproductive structures as separate organs, and anemochorous and anemophilous, i.e. pollen and seed are wind-dispersed. Low genetic diversity levels have been found in its natural range compared to other widespread conifers trees based on heterozygosity (H_e and H_o), nucleotide diversity (π) and effective population size (N_e) estimated using different types of markers (Yeh 1988; Shalev et al. 2022), which could potentially impact the species' long-term adaptability. However, other studies have found that largely distributed and outcrossing forest tree species, such as *Pinus pinea* or *Pinus halepensis*, despite their low levels of genetic diversity, may have an ecological or competitive advantage (Vendramin et al. 2008; Pinzauti et al. 2012).

While *T. plicata* is principally outcrossing, inbreeding by mating genetically related individuals or self-reproduction was reported in some situations (Russell et al. 2008; O'Connell et al. 2001; O'Connell et al. 2004) but inbreeding depression in growth and wood chemistry traits remained relatively low (Shalev et al. 2023; Wang and Russel 2006). Levels of genetic diversity vary along a latitudinal gradient within its natural distribution, showing higher levels in the south coastal region (O'Connell et al. 2008). This southern region was proposed as a refugium for *T. plicata* during the last glacial period and it later spread to the north and east to its current natural distribution (Yeh 1988; O'Connell et al. 2008). Predictions suggest potential further spread from coastal areas towards the interior due to changing climatic conditions (Hamann and Wang 2006).

The present study used Genotyping-by-Sequencing (GBS), a method for sequencing and genotyping that uses one or several restriction enzyme cutting sites (Andrews et al. 2016). GBS allows for large-scale direct discovery of single nucleotide polymorphisms (SNPs) that can be used in population genetic analysis without prior information. However, it requires deep data curation and complex bioinformatic pipelines (Vaux et al. 2023; Guillardín-Calvo et al. 2019). For example, estimating genotyping errors (Mastretta-Yanes et al. 2015) and missing data handling by imputation (Mora-Márquez et al. 2023) are key steps to process GBS data to obtain reliable results.

The genetic diversity of *T. plicata* woodlands in the UK may be affected by both the use of a restricted number of source provenances during their establishment as well as the inherently low genetic diversity reported in populations in the species' natural range. With more and more planted areas managed through natural regeneration, a lack of genetic diversity may be exacerbated over time depending on the level of gene flow between generations and across stands. We tested this hypothesis by analysing planted *T. plicata* adult trees and naturally regenerated juveniles from four different woodland sites in the UK. Our objectives were: (1) To perform SNP discovery in *T. plicata* by use of GBS with the selection of a subset of informative SNP markers for analysis; (2) To assess if the planted trees of each study site share genetically similar planting material origin and (3) To assess the genetic diversity in adults and juveniles across sites.

Materials and methods

Study sites

We sampled four planted woodland sites in England, UK (Fig. 1) that included *T. plicata* in the canopy, labelled here as Bagley Wood1, Longleat1, Stourhead1, and Stourhead2 (Table 1). The sites are being managed to encourage natural regeneration from the planted trees under a continuous cover forestry regime. Each site had at least 100 adult trees as recommended by Wojacki et al. (2019) to avoid biases

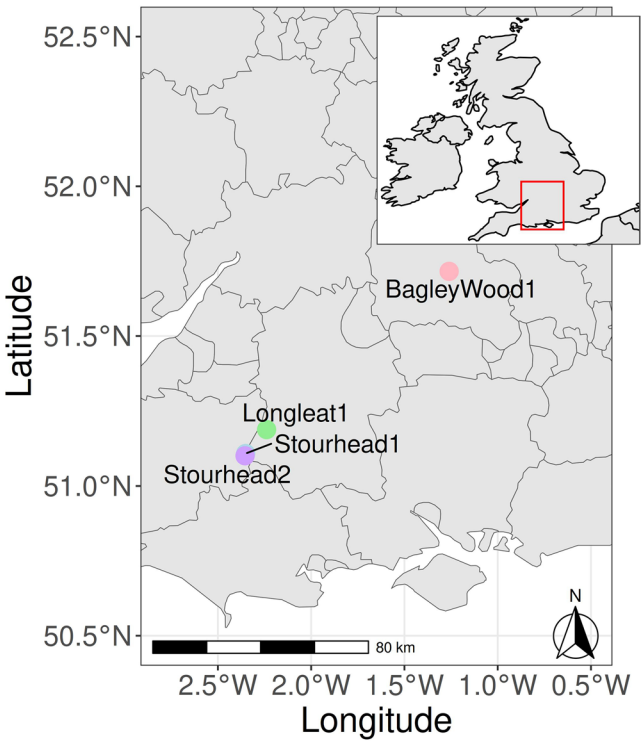


Fig. 1 Study sites location map

from sampling small populations, as well as having enough viable natural regeneration.

The plantations range from 63 to 74 years old, with Stourhead1 and Bagley Wood1 being the youngest and the oldest, respectively (Table 1). They are all privately owned (Table 1) and are situated in the south of England (Fig. 1). Bagley Wood1 is located the furthest (96 Km) from the other sites, while Longleat1 and Stourhead1 and 2 are at a 16 Km distance from each other. Stourhead2 (18.57 ha) is significantly larger than the other sites, which ranged from 3.73 to 5.07 ha (Table 1).

Genetic analysis

Sampling and DNA isolation

We sampled foliage from 10 adults and 10 juveniles across sites ($N=80$ in total). Adults were the original planted trees and juveniles were defined as naturally regenerated within the stand. Juvenile trees included seedlings and saplings from different age cohorts (maximum diameter basal height: 15 cm). The canopy trees were sampled by calculating random points using QGIS (QGIS 2023), resulting in a uniform distribution of both adults and juveniles across the site. The foliage samples were stored in paper envelopes and placed in plastic zip bags containing silica gel beads to promote their desiccation.

The DNA was isolated using the spin-column DNeasy plant mini kit (Qiagen USA, Valencia, CA) following the modified manufacturer's instructions (Guillardín and MacKay 2023). The DNA concentration was determined by using Qubit 4 Fluorometer (Invitrogen, Thermo Fisher Scientific, Cleveland, OH, USA). We used NanoDrop One Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE) to estimate the purity based on absorbance ratios A260/A280 and A260/A230. A minimum DNA concentration of 20 ng/µl and absorbance ratios of >1.8 were required for genotyping steps.

GBS library construction and sequencing

To select enzymes producing a suitable cut rate in the library preparation, we analysed the *T. plicata* genome sequence (Shalev et al. 2022) with the simulation software ddRADseqtools (Mora-Márquez et al. 2017). We compared the number and length of fragments recovered from each pair of available enzymes tested (Online resource 1) and selected the optimal combination (PstI/MspI).

The DNA samples were normalised to 20 ng/µl for library preparation (Genomic Analysis Platform of Laval University, Canada) and sequencing (Genome Quebec, Canada). The library was constructed with PstI/MspI enzyme combination (NEB, Ipswich, MA, USA) and the Illumina TruSeq HT Library Preparation Kit (Illumina, Inc. San Diego, CA,

Table 1 Study site details. Long: longitude, lat: latitude

Site	Year planted	Site area (ha)	Altitude (m)	Location	Ownership
Bagley Wood1	1950	3.73	93.1	Long: -1.260311 Lat: 51.715921	St John's College (Oxford)
Longleat1	1958	5.07	151	Long: -2.2389 Lat: 51.1885	Thynn family (Marquess of Bath)
Stourhead1	1961	5.00	244	Long: -2.35114 Lat: 51.10856	Hoare's family
Stourhead2	1951	18.57	244	Long: -2.35440 Lat: 51.10101	Hoare's family

USA), retaining the fragments above 300 bp from a fragment selection step. Sequencing was performed on an Illumina NovaSeq 6000 PE150 platform (Illumina Inc. San Diego, CA, USA), and 150 bp paired-end raw files were obtained. The fastq files were checked for quality using the FastQC software (Andrews 2010); the initial demultiplexing step used SABRE (Barry et al. 1989) based on the barcodes supplied by the Illumina sequencer provider, not allowing any mismatches or Ns, and the Illumina adapters were trimmed using Trimmomatic (Bolger et al. 2014). We trimmed the 3' end of the sequences and removed the adapter sequences.

Variant calling, filtering and genomic distribution of the SNPs

We used the IpyRAD pipeline (Eaton and Overcast 2020) to cluster, assemble and call the bases from the trimmed fastq files (Online resource 2). IpyRAD is divided into six steps (Fig. 2), allowing the repetition of steps separately once the results have been reviewed. After the first IpyRAD run including all the samples, we performed exploratory depth, missing data and number of loci per sample tests. We excluded five samples that showed an average depth below 6 reads or more than 50% of missing data. We reran IpyRAD from the 3rd step without the excluded samples. The reads were mapped to the reference genome (containing 67,899 scaffolds with a total size of 9.1 Gb) (Shalev et al. 2022) as shown in step 3 in Fig. 2. All the runs were performed in the University of Oxford Advanced Research Computing (Oxford, UK) facility and used a Capability cluster node with 48 CPUs and 384GB RAM.

First, the filtering of the initial GBS dataset was done to remove multi-allelic SNPs using PLINK (Purcell et al. 2007) and indels followed by a read-depth filtering between 6 and 300 and a minimum quality of 30 across the SNPs using VCFtools software (Danecek et al. 2011). Second, we pruned the SNPs under strong linkage disequilibrium by using PLINK (Purcell et al. 2007), calculating the pairwise correlation in a 50-SNP window and discarding SNPs with a correlation higher than 0.1, in increments of 10 SNPs at a time (Fig. 2).

Next, we used the *vcfR* package (Knaus and Grünwald 2017) loaded into RStudio (R Core Team 2018; RStudio Team 2020) to obtain a genlight object to perform the following population genetic analysis in the *DartR* (Mijangos et al. 2022) package environment. *DartR* facilitates the import and analysis of SNP data and includes the use of 3rd party population genetics software. We kept only the individuals with less than 40% of missing data. To ensure that the genetic variants are sufficiently common and informative, we used a cut-off of minor allele frequency (MAF) < 0.01 across the whole dataset. We also removed those SNPs that had a polymorphism information content (PIC) < 0.1 to retain highly informative SNPs. Finally, we determined the missing genotype data using the gtimputation pipeline which implements a machine learning procedure based on self-organising maps (SOM) (Mora-Márquez et al. 2023). We set up a 3 × 3 SOM dimension matrix, the minimum squared correlation (r2) of 0.1 and a maximum of five SNPs and we performed 1000 iterations (author's recommendation) (Mora-Márquez et al. 2023).

We used the *T. plicata* reference genome to determine the genomic distribution of SNPs. The *T. plicata* genome was assembled to a scaffold level; we used data on the size of scaffolds and numbers of SNPs per scaffold to calculate the SNP distribution at different stages of the filtering. We

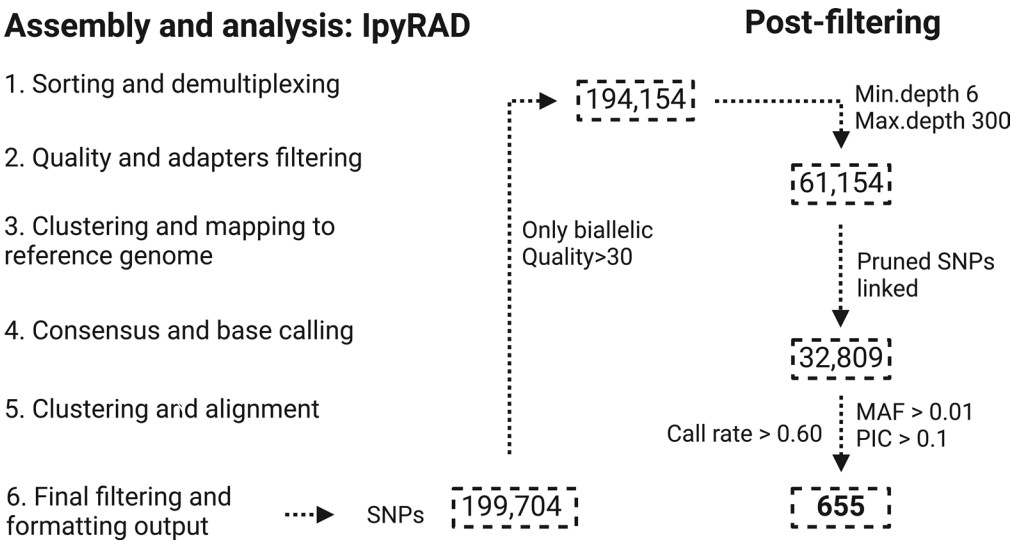


Fig. 2 Assembly and variant calling using IpyRAD and the following filtering steps of the recovered SNPs. MAF = minor allele frequency, PIC = polymorphism information content

calculated the distance among SNPs within scaffolds and only retained SNPs that were >1000 bp away from other SNPs (Savolainen and Pyhäjärvi 2007).

Genetic admixture, population structure and assignment analysis

We assessed the admixture and population structure using sparse nonnegative matrix factorization (sNMF) using the *LEA* R package (Frichot and François 2015). sNMF is computationally faster and statistically more robust than other likelihood-based methods such as STRUCTURE (Pritchard et al. 2000) and ADMIXTURE (Alexander et al. 2009), additionally, it is less affected by departures from population genetic assumptions (Frichot et al. 2014). We estimated the number of ancestral populations by testing K = 1–10 (rep = 10), and we selected K based on the cross-entropy criterion. We estimated and plotted each individual level of admixture for the selected number of ancestral populations found. We ran a Principal Coordinate Analysis (PCoA) (Zuur 2007) using *DartR* (Mijangos et al. 2022) to analyse and visualise the genetic variation and relationship among individuals.

The discriminant analysis of principal components (DAPC) (Jombart et al. 2010) allowed us to describe the diversity and structure of the sites and to assign the individuals to their genetic cluster without prior population information. We used the DAPC method implemented in the *adegenet* R package (Jombart et al. 2008) to find the number of genetic clusters (K) followed by an individual assignment analysis. We selected the optimal K following the Bayesian information criterion (BIC) statistical method (Jombart et al. 2010).

Genetic diversity

We estimated the genetic diversity within each site and compared adults and juveniles. We estimated observed heterozygosity (H_o), unbiased expected heterozygosity (uH_e), nucleotide diversity (π) and the inbreeding coefficient (FIS). Estimates of uH_e, H_o and FIS were made using the *DartR* package (Mijangos et al. 2022) for two scenarios, including and excluding private monomorphic SNPs following the recommendations of Schmidt et al. (2021). We determined the number and distribution of monomorphic SNPs across and within sites, and we used UpSetR (Conway et al. 2017) to visualize it. Significant differences among expected heterozygosity and FIS between adults and juveniles within sites were tested using *DartR* and the *diveRsity* (Keenan et al. 2013) R packages, respectively. FIS was calculated executing 1000 bootstraps to estimate the 95% confidence intervals and using Weir and Cockerham (1984) method.

We estimated the nucleotide diversity (π) across sites and for each site using the VCFtools software (Danecek et al. 2011). We estimated it in a 100 bp window and assessed its distribution per scaffold. We compared each group mean using the Kruskal-Wallis (Hollander et al. 2013) statistical test implemented in the *stats* package (R core Team 2018) followed by the Dunn-Bonferroni post-hoc test (Dunn 1964) using *FSA* R package (Ogle et al. 2023).

Results

Sequencing and variant calling

Simulations using ddRADseqTools (Mora-Márquez et al. 2017) and the reference genome (Shalev et al. 2022) suggested PstI/MspI as the optimal pair of restriction enzymes with 112,191 predicted DNA fragments. The enzymes ApeKI, BfaI and NsIi/MspI were estimated to produce > 300,000 fragments, and SbfI/MspI was estimated at < 10,000 fragments of 125–300 bp size, which appeared as too high and too low coverage of the *T. plicata* genome, respectively (Online resource 1). The FastQC analysis of the sequenced raw reads of all 80 individuals sampled gave high-quality scores with an average Phred quality score above 30.

After a first run of the assembler and base calling with the software IpyRAD (Eaton and Overcast 2020), three samples had read depth < 4 and two other samples had > 70% of missing data. The remaining 75 samples were retained for further data analysis, which included 1,273,904,871 raw reads (Online resource 3), and identified 194,154 biallelic SNPs variants called in 7,120 scaffolds (Fig. 2).

SNP data filtering and genomic distribution of the SNPs

The SNP-containing scaffolds varied in size and the majority were between 999 and 10,000 bp (Fig. 3, panel A). After filtering by depth, linkage disequilibrium, call rate, MAF and PIC, the SNPs retained were in 598 unique scaffolds, with a size ranging from 100,000 to 10,000,000 bp (Fig. 3, panel B and C). Additionally, the ratio of the number of scaffolds to the number of SNPs ranged from 0.04 before filtering to 0.91 after the filtering process. Most of the 598 unique scaffolds hold a single SNP (Fig. 3, panel D); however, 48 scaffolds had two SNPs separated by a mean distance of 805,500 bp (Table 2). Only three scaffolds had three SNPs and one scaffold had four SNPs, with an average of 1,474,563.5 and 1,965,387.0 bp, respectively (Table 2).

The set of 655 SNPs contained a significant but variable number of monomorphic SNPs in each population.

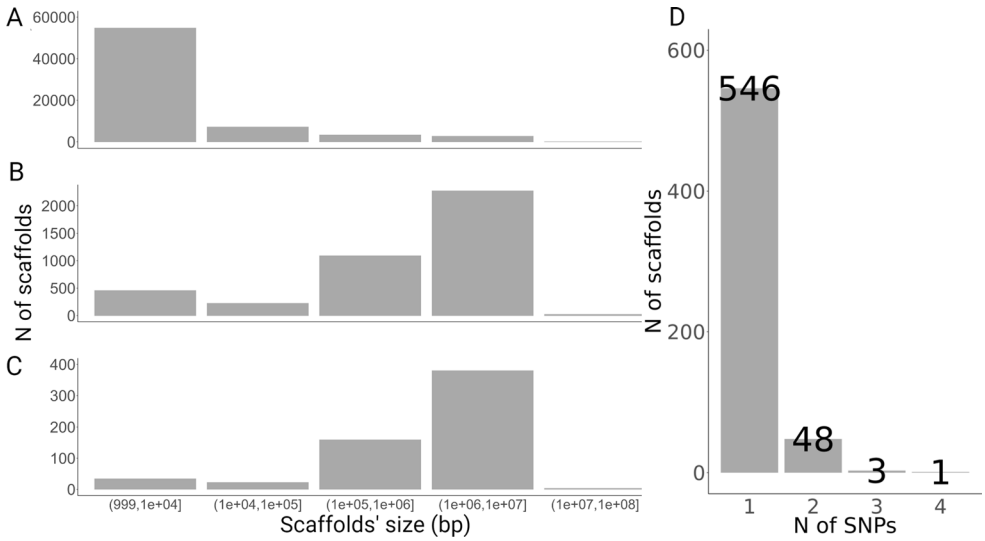


Fig. 3 Number of scaffolds and their size (bp) before filtering (**A**), after filtering by quality, depth and linkage disequilibrium (**B**) and after the final filtering of MAF, PIC and call rate yielding a total of 655 SNPs in 598 scaffolds (**C**) and number of SNPs per scaffolds (**D**)

Table 2 Physical distance (bp) between SNPs estimated for each of the scaffold type. N = number of scaffolds, 1st Qu. = 1st quartile, 3rd Qu. = 3rd quartile

Scaffold type	*N*	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2-SNPs	48	2,008	154,248.5	410,850.0	805,499.7	965,477.5	5,253,885
3-SNPs	3	484,188	839,187.8	876,692.5	1,474,563.5	1,766,007.5	3,720,676
4-SNPs	1	1,764,659	1,909,558.0	2,054,457.0	1,965,387.0	2,065,751.0	2,077,045

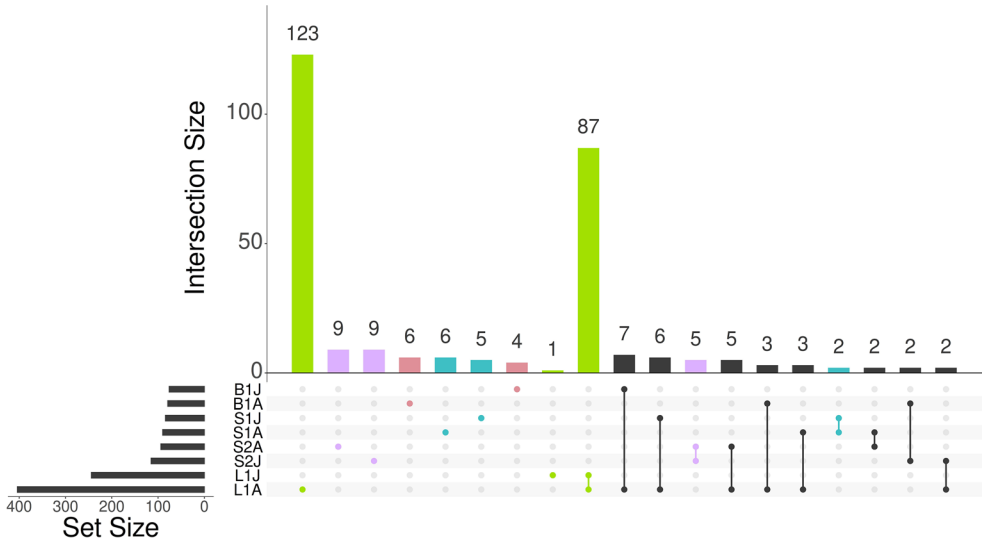


Fig. 4 Number of monomorphic SNPs distributed across sites (Bagley Wood1 (B1): red, Longleat1 (L1): green, Stourhead1 (S1): Blue, Stourhead2 (S2): Purple) and by group (adults (A) and juveniles (J). Set size is the total number of monomorphic SNPs in each site per group. The intersection size bars include the exclusive and shared monomorphic SNPs among and within sites identified by connected dots across the site and groups at the bottom of the figure

The distribution of monomorphic SNPs across sites was largely biased towards Longleat1 adults with 404 SNPs in total, representing > 60% all SNPs, of which 123 SNPs were exclusive (Fig. 4). The juveniles in Longleat1 had only one exclusive monomorphic SNP and 87 that overlapped with those found in the adults (Fig. 4). Both adults and juveniles in Bagley Wood1 had nine exclusive monomorphic SNPs, respectively, none of them shared within the site. Every other site had less than six unique monomorphic SNPs and shared at least two exclusive monomorphic SNPs between adults and juveniles.

no.

Population genetic analyses

Levels of admixture and number of genetic clusters

The minimum cross-entropy was found for three ancestral populations (0.631) (Fig. 5, panel A). Therefore, we reported the admixture coefficients for K = 3 (Fig. 5, panel B) and the respective q_1 (red), q_2 (green) and q_3 (blue) values (Online resource 4). Longleat1 (adults and juveniles) was composed of a single distinct ancestral population and showed no admixture ($q_3 \simeq 1$) in most of the individuals

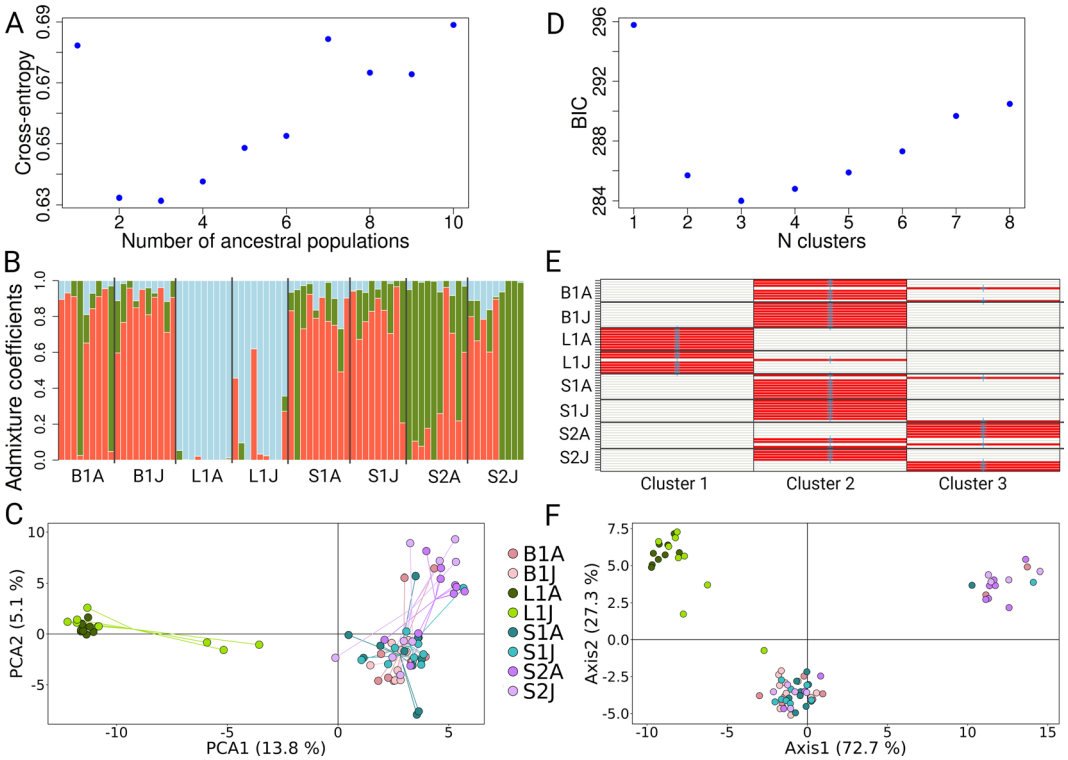


Fig. 5 Cross entropy plot for K = 1:10 (rep = 10) (**A**), admixture coefficients estimated for K = 3, q_1 (red), q_2 (green) and q_3 (blue), (**B**), principal coordinates analysis (PCoA) performed on the genotypes (**C**), Estimation of number of genetic clusters using BIC stat (K = 1:8) (**D**), individual assignment to each of the found genetic clusters, in red (**E**) and DAPC scatter plot

except for three juvenile trees exhibiting high levels of admixture ($q_3 < 7$ and $q_1 > 0.3$). Most Bagley Wood1 and Stourhead1 individuals showed low levels of admixture ($q_1 > 0.7$); however, admixture was found in some individuals from two ancestral populations revealing various levels of q_1 and q_2. Two adults from Bagley Wood1 showed high q_2 levels ($q_3 > 0.9$), indicating that their genetic composition is different from the other adults in that site. Stourhead2 trees showed high q_2 levels in most of the adults and juveniles, except for four adults and five juveniles which exhibited high q_1 values, indicating the presence of two different ancestral populations and their admixture in the site. This same pattern is shown in the PCoA (Fig. 5. panel C) where Longleat1 individuals were situated in an independent ancestry group, with 13.8% of variance explained by this coordinate. Stourhead2 appeared to have some individuals clustering alone (5.1% explained) while the rest were together with Bagley Wood1 and Longleat1 trees.

The DAPC analysis revealed three distinct genetic clusters across the four sites, estimated by the minimum BIC value (Fig. 5, panel D). The assignment of each individual to a genetic cluster showed a similar distribution to the sNMF results, except in Longleat1, where only one juvenile was in a different genetic cluster (Fig. 5, panel E and F).

Genetic diversity

The genetic diversity indicators (Table 3) were determined by using either the polymorphic SNPs alone or both the monomorphic and polymorphic SNPs together. Overall, we found low heterozygosity levels across and within populations. The observed heterozygosity (H_o) ranged from 0.125 to 0.213 across sites when using all makers, with the highest observed in Stourhead1 juveniles and, the lowest in Longleat1 adults. With the polymorphic SNPs alone, H_o generally increased across sites, with the highest H_o in Longleat1 adults (0.339) and the lowest in Stourhead1 juveniles (0.194). The largest change when excluding the monomorphic SNPs was observed in Longleat1 adults and juveniles (H_o from 0.125 to 0.339 and 0.146 to 0.325, respectively). We found significant differences in unbiased expected heterozygosity (uH_e) between adults and juveniles in Longleat1 (0.161 and 0.195) and Stourhead2 (0.275 and 0.289) using both monomorphic and polymorphic SNPs. However, we found significant but smaller differences between adults and juveniles at both in Longleat1 (0.437 and 0.420), Stourhead1 (0.333 and 0.353) when the monomorphic markers were excluded. The inbreeding coefficient (FIS) was less influenced by the markers used and ranged from 0.207 to 0.393 for both monomorphic and polymorphic SNPs and

Table 3 Genetic diversity parameters. Significant differences were tested between adults (A) and juveniles (J) within each site. Significant p-value < 0.05 is shown in the table as "*". N = number of samples collected, S.D = standard deviation, H_o = observed heterozygosity, uH_e = unbiased expected heterozygosity, FIS = inbreeding coefficient, N SNPs = number of polymorphic SNPs used to perform the analysis

	Bagley Wood1		Longleat1		Stourhead1		Stourhead2	
	A	J	A	J	A	J	A	J
N	9	10	9	9	10	9	10	9
N SNPs	655	655	655	655	655	655	655	655
Polymorphic SNPs	576	579	251	411	565	571	561	540
Monomorphic SNPs	79	76	404	244	90	84	94	115
Ho	0.211	0.204	0.125	0.146	0.159	0.213	0.180	0.154
HoSD	0.176	0.172	0.206	0.194	0.150	0.181	0.176	0.163
uHe	0.289	0.292	0.161	0.195*	0.275	0.289*	0.277	0.281
uHeSD	0.175	0.176	0.220	0.200	0.178	0.176	0.181	0.183
FIS	0.234	0.254	0.207	0.220	0.364	0.223	0.304	0.393
FISSD	0.387	0.384	0.419	0.420	0.405	0.400	0.418	0.439
Only polymorphic SNPs	510	510	232	232	496	496	476	476
Ho	0.254	0.240	0.339	0.325	0.194	0.260	0.228	0.197
HoSD	0.172	0.168	0.214	0.216	0.150	0.176	0.179	0.164
uHe	0.346	0.344	0.437*	0.420	0.333	0.353*	0.348	0.356
uHeSD	0.144	0.148	0.117	0.131	0.148	0.141	0.146	0.138
FIS	0.236	0.257	0.216	0.219	0.367	0.225	0.309	0.398
FISSD	0.387	0.384	0.427	0.429	0.401	0.398	0.417	0.433

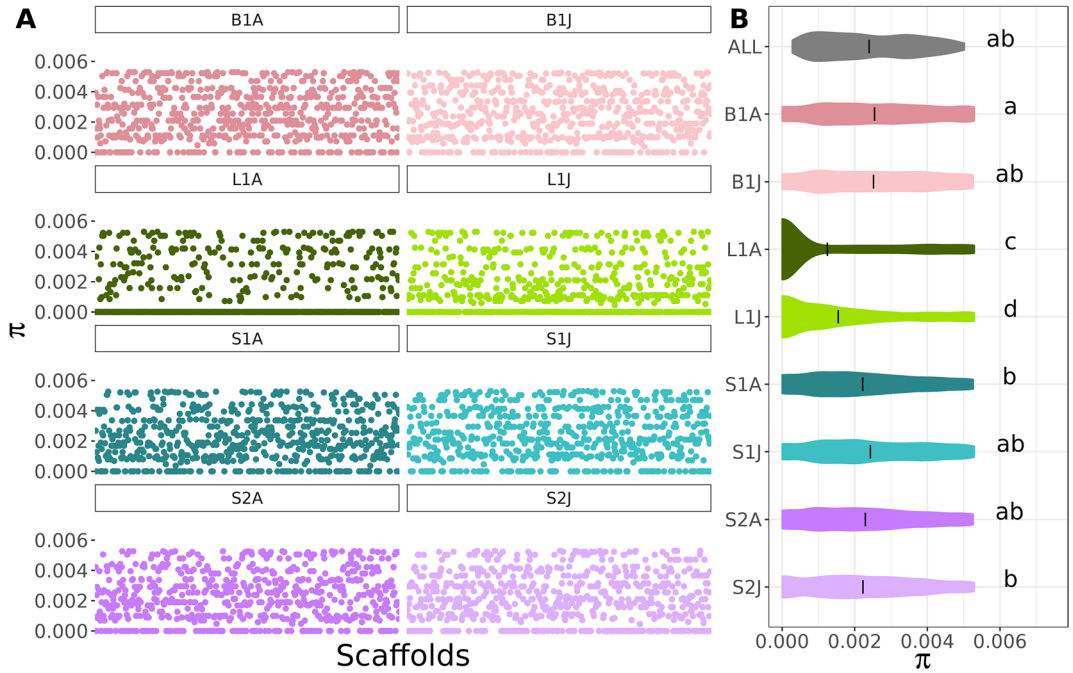


Fig. 6 Comparison of nucleotide diversity (π) distribution across sites estimated in each SNP (**A**) and the average per site and age cohort (**B**). Significant differences were tested using Kruskal-Wallis test (alpha = 0.05) between adults and juveniles across sites. Different letters designate significantly different groups tested by the post-hoc Dunn's test

0.216 to 0.398 for only polymorphic (Table 3). Longleat1 (both adults and juveniles) had the lowest inbreeding coefficient and Stourhead1 adults and Stourhead2 juveniles had the highest; however, no significant differences were found between adults and juveniles within sites.

An overall low nucleotide diversity was found across all the sites (π = 0.00226) and it ranged from 0.00136 in Longleat1 adults to 0.00263 in Bagley Wood1 (adults and juveniles). The distribution of the nucleotide diversity (π) per SNP within the scaffolds in each site is shown in Fig. 6 panel A. We detected a similar pattern across sites regarding the scaffolds containing SNPs with a high value of π. One population (Longleat1) gave significant differences in

nucleotide diversity between adults (0.00136) and juveniles (0.00167) (Fig. 6 panel B).

Discussion

In this study, we assessed population structure and genetic diversity in planted *T. plicata* trees and their offspring in four UK woodlands. We identified 655 informative SNP markers by using a GBS approach. We found three distinct genetic clusters across the four sites with low levels of admixture in the adult and juvenile trees. These clusters aligned with significant differences in genetic variation and in the number of private monomorphic SNPs across sites. Overall, the heterozygosity levels were low (uH_e and H_o) across and within sites. Two sites (Longleat1 and Stourhead1) showed significantly higher uH_e in the juvenile trees compared to adults. Additionally, we found very low nucleotide diversity π across and within sites, with adults ranging from 0.00125 to 0.002551, and juveniles slightly higher at 0.00155 to 0.00252, which may indicate low levels of genetic diversity and therefore low adaptation potential.

Distribution of variants across the genome and monomorphic SNPs

We found 199,704 variants before filtering, which is relatively low compared to the 1.3 M SNPs found in Chinese fir (*Cunninghamia lanceolata*) and 3 M SNPs found in *Thuja sutchuenensis* (Tao et al. 2024) also both members of the Cupressaceae (Zheng et al. 2019), and the 349,542 SNPs found in the foxtail pine (*Pinus balfouriana*) (Friedline et al. 2015). We retained 655 high-quality SNPs with a high proportion of private monomorphic SNPs mostly found in Longleat1 (Fig. 4). The MAF > 0.01 filter usually removes rare, private and monomorphic SNPs across the whole dataset. In our case, the majority of the site-specific monomorphic SNPs are found in Longleat1, suggesting differences in the provenance of the seed sources used to establish the four plantations. A substantial difference in the Longleat1 private monomorphic alleles were found between adults (123) and juveniles (1). However, 87 monomorphic SNPs were shared by both groups, which may reveal higher genetic variation in the naturally regenerated offspring than in the planted adults. The low genetic variation found in the adults may suggest a high level of fixed alleles compared to the other sites.

The majority of the 598 SNP-containing scaffolds we identified in *T. plicata* were over 100 Kb in size and were in the top 9% longest scaffolds in the genome assembly (Shalev et al. 2022). Conifer tree species have large genomes and most have been assembled to a scaffold level, such as *Picea sitchensis* (Gagalova et al. 2022), *Pinus taeda* (Zimin et al. 2014) or *Pseudotsuga menziesii* (Neale et al. 2017). Angiosperm tree species have shorter genomes and several have been assembled to a chromosome level including *Quercus robur* (Plomion et al. 2016) and *Fraxinus excelsior* (Sollars et al. 2017). Scaffolds are aligned sequences (contigs) that form larger regions. Due to their size, the position of scaffolds within conifer genomes is usually not known, meaning that the relative genomic or chromosomal location of variants (e.g., SNPs) found across different scaffolds cannot be directly determined (Luo et al. 2021). However, a relatively uniform distribution of SNPs along the genome is essential for an optimal estimation of population genetics parameters (Brhane et al. 2022). One way to overcome this problem is to limit the number and proximity of SNPs retained per scaffold, which is the approach followed in the present report. Other studies followed a similar SNP filtering process to perform population genetics analysis, for example in a sea fly where 93.8% of the retained SNPs were included in the largest scaffolds (> 100 Kb) in the reference genome (Muñoz et al. 2016).

Population genetic analysis

Three ancestral populations were found suggesting that the study sites were established using plant material from three different origins. Every site showed individuals with some level of admixture between two genetic clusters except for Longleat1 adults which had no admixture. However, three juveniles in Longleat1 had high levels of admixture indicating gene flow from an external source, although the presence of non-characterised genetically distinct adults in the stand cannot be ruled out. Bagley Wood1 and Stouhead1 shared most of their genetic makeup while Stourhead2 included some individuals with admixed backgrounds; however, the majority were assigned to two different ancestral populations. This may suggest that Bagley Wood1 and Stourhead1 were established using an identical source while sites within the same woodland (Stourhead1 and 2) were established using different plant materials.

Overall, H_o levels across sites were relatively low (ranging from 0.125 to 0.213) but followed the low level (H_o = 0.219) found in the natural range of *T. Plicata* (Shalev et al. 2022). Due to a lack of population structure and genetic differentiation across its large and continuous natural distribution, it has been suggested that *T. plicata* experienced an historical bottleneck followed by a range expansion that reduced its genetic variation, (Shalev et al. 2022). However, despite its low genetic diversity, *T. plicata* may hold a high potential for adaptation and further expansion of its natural range as discussed by these authors. Similarly, *Pinus pinea* has low genetic diversity (H = 0.11, from nuclear

microsatellites) (Pinzauti et al. 2012) across its large natural distribution and is suggested to hold certain competitive advantage, potentially due to a high level of local adaptation (Simões et al. 2024). However, exotic planted woodlands that are managed using natural regeneration may be affected by the initial low levels of genetic diversity of the seed source. This may be followed by a lack of gene flow due to isolation from other populations, a reduced number of breeding parents due to stand thinning or logging, or selective mating may decrease the potential for adaptation to climate change in future generations.

Other studies following a similar approach as used here showed higher heterozygosity levels in *Pseudotsuga menziesii* stands (H_o = 0.291) (Galeano et al., 2021) and *Pinus strobus* second growth forests (H_o = 0.232) (Rajora and Zinck, 2021). Low levels of H_o in Longleat1 is explained by the large number of monomorphic SNPs, which could suggest the potential fixation of alleles or high levels of inbreeding in the source populations. The other sites showed low H_o despite the lower proportion of monomorphic SNPs, which may indicate low genetic diversity and may lead to potential lack of adaptability to current and future changes in the climate. Our FIS results (ranging from 0.207 to 0.393) showed similar inbreeding in these planted woodlands than in the natural range (FIS = 0.331) (Shalev et al. 2022). The observed heterozygosity (H_o) was similar between adults and juveniles produced by natural regeneration across all sites. This may indicate that the frequency of heterozygotes on each site has not been reduced suggesting an efficient transfer of alleles to the next generation of trees. These results follow the observations of Liesebach et al. (2024) in *Picea abies* and *Fagus sylvatica* stands in Europe, where H_o was consistent between adults and naturally regenerated juveniles.

The overall estimated nucleotide diversity (π) we found in *T. plicata* was 0.00226, which is in the same range as in the natural distribution of the species (0.0020) (Shalev et al. 2022). These levels of diversity are relatively low compared to estimates in *Picea abies* (0.0049 to 0.00625) (Wang et al. 2020) or *Cupressus chengiana* (0.0077) (Li et al. 2020) but are higher than found in the congeneric *Thuja sutchuenensis* (0.00186) (Tao et al. 2024). *Thuja sutchuensis* is an endangered species and its low π may be due to the fragmentation and the small size of existing populations. The low genetic diversity found in UK plantations of *T. plicata* most likely reflects the levels of genetic drift and inbreeding in the North-American populations used for their establishment (Tao et al. 2024).

Practical application and conclusions

We identified a set of markers to evaluate genetic variation and genetic diversity in *T. plicata* populations which could be useful in further investigations of populations and genetic resources. We found many monomorphic SNPs unique to the adults from Longleat1, indicative of a lower genetic variation within this site and a distinct genetic background. Three genetic clusters were found across the four woodland sites. One cluster was unique for Longleat1 while one was assigned to the majority of individuals in Bagley Wood1 and Stourhead1, suggesting the stands are comprised of materials from similar origin, and the third genetic cluster included individuals from Stourhead2.

The consistently low heterozygosity and nucleotide diversity levels in adults and juveniles in Longleat1, together with the assignment of all the individuals to a unique cluster suggests that the Longleat1 germplasm may come from an area of lower genetic diversity. However, the admixture found in some of the juveniles may indicate that gene flow from external populations is widening the gene pool in the site. These observations indicate that the current genetic diversity is maintained and even increases in some sites across generations. However, although the inherently low genetic diversity in *T. plicata* does not appear to be maladaptive in its natural range, its occurrence in relatively isolated woodlands opens the door to inbreeding and genetic drift, with potential impacts on genetic diversity and adaptability to to changing environmental conditions.

Our study combined analyses of population structure and genetic diversity in four *T. plicata* populations in the UK thus producing integrated insights for woodland management. Future studies may aim to improve our understanding of the species as a non-native, e.g. by mapping the origin of UK planted materials to areas in its natural range to uncover the population genetics of monomorphic alleles only observed in Longleat1 adults. Our results suggest that a potential injection of genetic diversity (e.g. through planting) may be needed in woodlands established from a narrow gene pool in order to maintain or increase genetic diversity (heterozygosity and nucleotide diversity).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11295-025-01689-7.

Acknowledgements We want to acknowledge the forestry estates where the samples came from: Stourhead (Western), Longleat (Wiltshire) and Bagley Wood. We also thank Barley Rose Collier Harris for helping with sampling, and Will Hoare, George MacKay and Emily Farley for their assistance in both laboratory and field work. We would like to thank Dr Hayley Tumas for reviewing the paper. We also want to thank the three anonymous reviewers and the associate editor that improved the final version of this manuscript. Finally, we would like to acknowledge the use of the University of Oxford Advanced Research

Computing (ARC) facility in carrying out this work. https://doi.org/10.5281/zenodo.22558.

Author contributions LG drafted the manuscript and both LG and JM revised the main manuscript. LG conceived and planned the experiments, performed laboratory and statistical analyses and prepared the figures. LG and EG performed the sampling and developed the field survey. JM supervised the project.

Funding LG received financial support from the Oxford-John Oldacre Graduate Scholarship and from Mr Henry Hoare. The project was funded in part by the OxLEP fund for the Oxford Plant Sciences Centre for Innovation.

Data availability The raw NGS read data was deposited to NCBI's Short Read Archive (SRA). The BioProject ID is PRJNA1081148 and the SRA accession numbers range from SAMN40149401 to SAMN40149480. The SNP set generated and analyzed during the current study will be public available in the open access repository Github (https://github.com/LauGuillardin/GenomicExploration_Thujaplicata).

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Alberto FJ, Aitken SN, Alía R, González-Martínez SC, Hänninen H, Kremer A, Lefèvre F, Lenormand T, Yeaman S, Whetten R, Savolainen O (2013) Potential for evolutionary responses to climate change – evidence from tree populations. Glob Change Biol 19(6):1645–1661. https://doi.org/10.1111/GCB.12181

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19(9):1655–1664. https://doi.org/10.1101/gr.094052.109

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet 17(2) Article 2. https://doi.org/10.1038/nrg.2015.28

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom

Baldi P, Porta N (2022) Toward the genetic improvement of Drought Tolerance in conifers: an Integrated Approach. Forests. https://doi.org/10.3390/f13122016

Barry JT, Francis BJ, Davies RB (1989) SABRE: Software for the analysis of Binary recurrent events. In: Decarli A, Francis BJ, Gilchrist R, Seeber GUH (eds) Statistical modelling. Lecture Notes in Statistics, vol 57. Springer, New York, pp 56–63. https://doi.org/10.1007/978-1-4612-3680-1_7

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brhane H, Haileselassie T, Tesfaye K, Ortiz R, Hammenhag C, Abreha KB, Geleta M (2022) Novel GBS-Based SNP markers for Finger Millet and their use in genetic diversity analyses. Front Genet 13. https://doi.org/10.3389/fgene.2022.848627

Conway JR, Lex A, Gehlenborg N (2017) UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 33(18):2938–2940. https://doi.org/10.1093/bioinformatics/btx364

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. Bioinformatics 27(15):2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dunn OJ (1964) Multiple comparisons using rank sums. Technometrics 6(3):241–252. https://doi.org/10.1080/00401706.1964.10490181

Eaton DAR, Overcast I (2020) IpyRAD: interactive assembly and analysis of RADseq datasets. Bioinformatics 36(8):2592–2594. https://doi.org/10.1093/bioinformatics/btz966

Forest Research (2022) National Forest Inventory: 25-year forecast of softwood timber availability

Frichot E, François O (2015) LEA: an R package for landscape and ecological association studies. Methods Ecol Evol 6(8):925–929. https://doi.org/10.1111/2041-210X.12382

Frichot E, Mathieu F, Trouillon T, Bouchard G, François O (2014) Fast and efficient estimation of individual ancestry coefficients. Genetics 196(4):973. https://doi.org/10.1534/genetics.113.160572

Friedline CJ, Lind BM, Hobson EM, Harwood DE, Mix AD, Maloney PE, Eckert AJ (2015) The genetic architecture of local adaptation I: the genomic landscape of foxtail pine (Pinus balfouriana Grev. And Balf.) As revealed from a high-density linkage map. Tree Genet Genomes 11(3):49. https://doi.org/10.1007/s11295-015-0866-x

Gagalova KK, Warren RL, Coombe L, Wong J, Nip KM, Yuen MMS, Whitehill JGA, Celedon JM, Ritland C, Taylor GA, Cheng D, Plettner P, Hammond SA, Mohamadi H, Zhao Y, Moore RA, Mungall AJ, Boyle B, Laroche J, Cottrell J, Mackay JJ, Lamothe M, Gérardi S, Isabel N, Pavy N, Jones SJM, Bohlmann J, Bousquet J, Birol I (2022) Spruce giga-genomes: structurally similar yet distinctive with differentially expanding gene families and rapidly evolving genes. Plant J 111(5):1469–1485. https://doi.org/10.1111/tpj.15889

Galeano E, Bousquet J, Thomas BR (2021) SNP-based analysis reveals unexpected features of genetic diversity, parental contributions and pollen contamination in a white spruce breeding program. Sci Rep 11:4990. https://doi.org/10.1038/s41598-021-84566-2

Gil-Moreno D, Ridley-Ellis D, McLean P (2016) Timber properties of noble fir, Norway spruce, western red cedar and western hemlock grown in Great Britain. Forestry Commission library review

Guillardín-Calvo L, Mora-Márquez F, Soto Á, López de Heredia U (2019) RADdesigner: a workflow to select the optimal sequencing methodology in genotyping experiments on woody plant species. Tree Genet Genomes 15:64. https://doi.org/10.1007/s11295-019-1372-3

Guillardín L, MacKay JJ (2023) Comparing DNA isolation methods for forest trees: quality, plastic footprint, and time-efficiency. Plant Methods 19:111. https://doi.org/10.1186/s13007-023-01086-y

Hamann A, Wang T (2006) Potential effects of climate change on ecosystem and tree species distribution in British Columbia. Ecology

87(11):2773–2786. https://doi.org/10.1890/0012-9658(2006)87[2773:peocco]2.0.co;2
Hebda RJ, Mathewes RW (1984) Holocene History of Cedar and native Indian cultures of the North American Pacific Coast. Science 225(4663):711–713. https://doi.org/10.1126/science.225.4663.711
Heuertz M, Carvalho SB, Galindo J, Rinkevich B, Robakowski P, Aavik T, Altinok I, Barth JMI, Cotrim H, Goessen R, González-Martínez SC, Grebenc T, Hoban S, Kopatz A, McMahon BJ, Porth I, Raeymaekers JAM, Träger S, Valdecantos A et al (2023) The application gap: Genomics for biodiversity and ecosystem service management. Biol Conserv 278:109883. https://doi.org/10.1016/J.BIOCON.2022.109883
Hoban S, Bruford M, D'Urban Jackson J, Lopes-Fernandes M, Heuertz M, Hohenlohe PA, Paz-Vinas I, Sjögren-Gulve P, Segelbacher G, Vernesi C, Aitken SN, Bertola LD, Bloomer P, Breed M, Rodríguez-Correa H, Funk WC, Grueber CE, Hunter ME, Jaffe R et al (2020) Genetic diversity targets and indicators in the CBD post-2020 global Biodiversity Framework must be improved. Biol Conserv 248:108654. https://doi.org/10.1016/j.biocon.2020.108654
Hoban S, Bruford MW, da Silva JM, Funk WC, Frankham R, Gill MJ, Grueber CE, Heuertz M, Hunter ME, Kershaw F, Lacy RC, Lees C, Lopes-Fernandes M, MacDonald AJ, Mastretta-Yanes A, McGowan PJK, Meek MH, Mergeay J, Millette KL et al (2023) Genetic diversity goals and targets have improved, but remain insufficient for clear implementation of the post-2020 global biodiversity framework. Conserv Genet. https://doi.org/10.1007/s10592-022-01492-0
Hollander M, Wolfe DA, Chicken E (2013) Nonparametric statistical methods. Wiley
Jombart (2008) Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24(11):1403–1405. https://doi.org/10.1093/bioinformatics/btn129
Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:94. https://doi.org/10.1186/1471-2156-11-94
Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA (2013) diveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. Methods Ecol Evol 4(8):782–788. https://doi.org/10.1111/2041-210X.12067
Kerr G (2019) Summary of FR seed origin trials on western red cedar (Thuja plicata D. Don). Forestry Commission Research Note
Knaus BJ, Grünwald NJ (2017) Vcfr: a package to manipulate and visualize variant call format data in R. Mol Ecol Resour 17(1):44–53. https://doi.org/10.1111/1755-0998.12549
Li J, Milne RI, Ru D, Miao J, Tao W, Zhang L, Xu J, Liu J, Mao K (2020) Allopatric divergence and hybridization within Cupressus Chengiana (Cupressaceae), a threatened conifer in the northern Hengduan Mountains of western China. Mol Ecol 29(7):1250–1266. https://doi.org/10.1111/mec.15407
Liesebach H, Eusemann P, Höltken AM, Tröber U, Kuchma O, Karopka M, Becker F, Kätzel R, Fussi B (2024) Effective population size of adult and offspring cohorts as a genetic monitoring tool in two stand-forming and wind-pollinated tree species: Fagus sylvatica L. and Picea abies (L.) Karst. Conserv Genet 1572–9737. https://doi.org/10.1007/s10592-024-01600-2
Luo J, Wei Y, Lyu M, Wu Z, Liu X, Luo H, Yan C (2021) A comprehensive review of scaffolding methods in genome assembly. Brief Bioinform 22(5):bbab033. https://doi.org/10.1093/bib/bbab033
Mastretta-Yanes A, Arrigo N, Alvarez N, Jorgensen TH, Piñero D, Emerson BC (2015) Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. Mol Ecol Resour 15(1):28–41. https://doi.org/10.1111/1755-0998.12291
Mijangos JL, Gruber B, Berry O, Pacioni C, Georges A (2022) dartR v2: an accessible genetic analysis platform for conservation, ecology and agriculture. Methods Ecol Evol 13(10):2150–2158. https://doi.org/10.1111/2041-210X.13918
Mora-Márquez F, García-Olivares V, Emerson BC, López de Heredia U (2017) Ddradseqtools: a software package for in silico simulation and testing of double-digest RADseq experiments. Mol Ecol Resour 17(2):230–246. https://doi.org/10.1111/1755-0998.12550
Mora-Márquez F, Nuño JC, Soto Á, López de Heredia U (2023) Missing genotype imputation in non-model species using Self-Organizing maps [Preprint]. https://doi.org/10.22541/au.168596316.62218099/v1. Preprints
Muñoz J, Chaturvedi A, De Meester L, Weider LJ (2016) Characterization of genome-wide SNPs for the water flea Daphnia pulicaria generated by genotyping-by-sequencing (GBS). Sci Rep 6(1):28569. https://doi.org/10.1038/srep28569
Neale DB, McGuire PE, Wheeler NC, Stevens KA, Crepeau MW, Cardeno C, Zimin AV, Puiu D, Pertea GM, Sezen UU, Casola C, Koralewski TE, Paul R, Gonzalez-Ibeas D, Zaman S, Cronn R, Yandell M, Holt C, Langley CH, Yorke JA, Salzberg SL, Wegrzyn JL (2017) The Douglas-Fir genome sequence reveals specialization of the photosynthetic apparatus in Pinaceae. Genes Genomes Genet G3(9):3157–3167. https://doi.org/10.1534/g3.117.300078
O'Connell LM, Ritland K, Thompson SL (2008) Patterns of post-glacial colonization by western redcedar (Thuja plicata, Cupressaceae) as revealed by microsatellite markers. Botany 86(2):194–203. https://doi.org/10.1139/B07-124
O'Connell LM, Russell J, Ritland K (2004) Fine-scale estimation of outcrossing in western redcedar with microsatellite assay of bulked DNA. Heredity 93(5):443–449. https://doi.org/10.1038/sj.hdy.6800521
O'Connell LM, Viard F, Russell J, Ritland K (2001) The mating system in natural populations of western redcedar (Thuja plicata). Can J Bot 79(6):753–756. https://doi.org/10.1139/cjb-79-6-753
Ogle DH, Doll JC, Wheeler AP, Dinno A (2023) FSA: Simple Fisheries Stock Assessment Methods. R package version 0.9.5. https://fish r-core-team.github.io/FSA/
Pierro E, Mosca E, González-Martínez S, Binelli G, Neale D, Porta N (2017) Adaptive variation in natural Alpine populations of Norway spruce (*Picea abies* [L.] Karst) at regional scale: Landscape features and altitudinal gradient effects. For Ecol Manag 405:350–359. https://doi.org/10.1016/J.FORECO.2017.09.045
Pinzauti F, Sebastiani F, Budde KB, Fady B, González-Martínez SC, VendraminGG (2012) Nuclear microsatellites for *Pinus pinea* (Pinaceae), a genetically depauperate tree, and their transferability to *P. Halepensis*. Am J Bot 99(9). https://doi.org/10.3732/ajb.1200064
Plomion C, Aury JM, Amselem J, Alaeitabar T, Barbe V, Belser C, Bergès H, Bodénès C, Boudet N, Boury C, Canaguier A, Couloux A, Da Silva C, Duplessis S, Ehrenmann F, Estrada-Mairey B, Fouteau S, Francillonne N, Gaspin C, Guichard C, Klopp C, Labadie K, Lalanne C, Le Clainche I, Leplé JC, Le Provost G, Leroy T, Lesur I, Martin F, Mercier J, Michotey C, Murat F, Salin F, Steinbach D, Faivre-Rampant P, Wincker P, Salse J, Quesneville H, Kremer A (2016) Decoding the oak genome: public release of sequence data, assembly, annotation and publication strategies. Mol Ecol Resour 16(1):254–265. https://doi.org/10.1111/1755-0998.12425
Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population structure using multilocus genotype data. Genetics 155(2):945–959. https://doi.org/10.1093/genetics/155.2.945
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007) PLINK: a Tool Set for whole-genome Association and

Population-based linkage analyses. Am J Hum Genet 81(3):559–575. https://doi.org/10.1086/519795

QGIS (2023) QGIS Geographic Information System. Open Source Geospatial Foundation Project. (3.32.3) [Computer software]. http://qgis.org

R Core Team (2018) R: a language and environment for statistical computing (4.3.0) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rajora OP, Zinck JWR (2021) Genetic diversity, structure and effective population size of old-growth vs. second-growth populations of keystone and long-lived conifer, Eastern White Pine (*Pinus strobus*): conservation value and climate adaptation potential. Front Genet 12. https://doi.org/10.3389/fgene.2021.650299

RStudio Team (2020) RStudio: Integrated development for R (2023.06.0) [Computer software]. Posit. http://www.rstudio.com/

Russell JH, Ferguson DC (2008) Preliminary results from five generations of a western redcedar (Thuja plicata) selection study with self-mating. Tree Genet Genomes 4(3):509–518. https://doi.org/10.1007/s11295-007-0127-8

Savolainen O, Pyhäjärvi T (2007) Genomic diversity in forest trees. Curr Opin Plant Biol 10(2):162–167. https://doi.org/10.1016/j.pbi.2007.01.011

Schmidt TL, Jasper M-E, Weeks AR, Hoffmann AA (2021) Unbiased population heterozygosity estimates from genome-wide sequence data. Methods Ecol Evol 12(10):1888–1898. https://doi.org/10.1111/2041-210X.13659

Sgrò CM, Lowe AJ, Hoffmann AA (2011) Building evolutionary resilience for conserving biodiversity under climate change. Evol Appl 4(2):326–337. https://doi.org/10.1111/j.1752-4571.2010.00157.x

Shalev TJ, Gamal El-Dien O, Yuen MMS, Shengqiang S, Jackman SD, Warren RL, van der Coombe L, Stewart A, Boston LB, Plott C, Jenkins J, He G, Yan J, Yan M, Guo J, Breinholt JW, Neves LG, Grimwood J et al (2022) The western redcedar genome reveals low genetic diversity in a self-compatible conifer. Genome Research 1952–1964. https://doi.org/10.1101/gr.276358.121

Shalev TJ, Gamal El-Dien O, Yuen MMS, van der Merwe L, Kirst M, Yanchuk AD, Ritland C, Russell JH, Bohlmann J (2023) Genetic architecture of terpene chemistry and growth traits and the impact of inbreeding on these traits in western redcedar (Thuja plicata). Evol Appl. https://doi.org/10.1111/EVA.13526

Simões ASB, Borges MM, Grazina L, Nunes J (2024) Stone Pine (*Pinus pinea* L.) High-added-Value Genetics: an overview. Genes 15(1):84. https://doi.org/10.3390/genes15010084

Sollars ESA, Harper AL, Kelly LJ, Sambles CM, Ramirez-Gonzalez RH, Swarbreck D, Kaithakottil G, Cooper ED, Uauy C, Havlickova L, Worswick G, Studholme DJ, Zohren J, Salmon DL, Clavijo BJ, Li Y, He Z, Fellgett A, McKinney LV, Nielsen LR, Douglas GC, Kjær ED, Downie JA, Boshier D, Lee S, Clark J, Grant M, Bancroft I, Caccamo M, Richard J, Buggs RJA (2017) Genome sequence and genetic diversity of European ash trees. Nature 541(7636):212–216. https://doi.org/10.1038/nature20786

Stirling R, Sturrock RN, Braybrooks A (2017) Fungal decay of western redcedar wood products – a review. Int Biodeterior Biodegradation 125:105–115. https://doi.org/10.1016/j.ibiod.2017.09.001

Tao T, Milne RI, Li J, Yang H, Wang S, Chen S, Mao K (2024) Conservation genomic investigation of an endangered conifer, Thuja sutchuenensis, reveals low genetic diversity but also low genetic load. Plant Divers 46(1):78–90. https://doi.org/10.1016/j.pld.2023.06.005

Vaux F, Dutoit L, Fraser CI, Waters JM (2023) Genotyping-by-sequencing for biogeography. J Biogeogr 50(2):262–281. https://doi.org/10.1111/jbi.14516

Vendramin GG, Fady B, González-Martínez SC, Hu FS, Scotti I, Sebastiani F, Soto Á, Petit RJ (2008) Genetically depauperate but widespread: the case of an emblematic Mediterranean pine. Evolution 62(3):680–688. https://doi.org/10.1111/j.1558-5646.2007.00294.x

Wang T, Russel JH (2006) Evaluation of selfing effects on western redcedar growth and yield in operational plantations using the tree and stand simulator (TASS). For Sci 52(3):281–289. https://doi.org/10.5849/forsci.15-042

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 Biol Evol 12(2):3803–3817. https://doi.org/10.1093/gbe/evaa005

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the analysis of Population structure. Evolution 38(6):1358. https://doi.org/10.2307/2408641

Wernberg T, Coleman MA, Bennett S, Thomsen MS, Tuya F, Kelaher BP (2018) Genetic diversity and kelp forest vulnerability to climatic stress. Sci Rep 8(1):1851. https://doi.org/10.1038/s41598-018-20009-9

Wojacki J, Eusemann P, Ahnert D, Pakull B, Liesebach H (2019) Genetic diversity in seeds produced in artificial Douglas-fir (Pseudotsuga menziesii) stands of different size. For Ecol Manag 438(February):18–24. https://doi.org/10.1016/j.foreco.2019.02.012

Yeh FC (1988) Isozyme variation of Thuja plicata (cupressaceae) in British Columbia. Biochem Syst Ecol 16(4):373–377. https://doi.org/10.1016/0305-1978(88)90028-2

Zheng H, Hu D, Wei R, Yan S, Wang R (2019) Chinese fir breeding in the high-throughput sequencing era: insights from SNPs. Forests 10(8) Article 8. https://doi.org/10.3390/f10080681

Zimin A, Stevens KA, Crepeau MW, Holtz-Morris A, Koriabine M, Marçais G, Puiu D, Roberts M, Wegrzyn JL, de Jong PJ, Neale DB, Salzberg SL, Yorke JA, Langley CH (2014) Sequencing and assembly of the 22-gb loblolly pine genome. Genetics 196(3):875–890. https://doi.org/10.1534/genetics.113.159715

Zuur (2007) Principal coordinate analysis and non-metric multidimensional scaling. In Analysing Ecological Data (pp. 259–264). Springer New York. https://doi.org/10.1007/978-0-387-45972-1_15

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.