

RESEARCH

Open Access



# Molecular characterisation of *Pinus sylvestris* (L.) in Ireland at the western limit of the species distribution

Samuel Belton<sup>1†</sup>, Philippe Cubry<sup>1,2†</sup>, Jenni R. Roche<sup>3</sup> and Colin T. Kelleher<sup>1\*</sup>

## Abstract

**Background** Scots pine (*Pinus sylvestris* L.) underwent significant population declines across much of northwest Europe during the mid-to-late Holocene and was thought to have become extirpated in Ireland from about 400 AD. However, most extant populations are plantations reintroduced from Scotland. Others are naturalised therefrom and one in Western Ireland is a putative relict. In this paper, Scots pine in Ireland are genetically described for the first time.

**Results** Using two mitochondrial (mtDNA) loci, eight chloroplast (cpSSR) and 18 nuclear (nSSR) loci, the genetic composition and diversity of 19 Irish Scots pine populations is described and compared to other European populations. All trees sampled in Ireland were fixed for mitotype *a*, which is the most common across northwest Europe. By contrast, cpSSR ( $H_{CP} = 0.967$ ) and nSSR ( $H_e = 0.540$ ) variation was high, and comparable with estimates for other regions across the species range. Differentiation at both sets of loci were similarly low (cpSSR  $F_{ST} = 0.019$ ; nSSR  $F_{ST} = 0.018$ ), but populations from continental Europe were significantly differentiated from all Irish populations based on nSSR variation.

**Conclusions** All Irish Scots pine are likely part of a common Irish-Scottish gene pool which diverged from continental Scots pine following post-glacial recolonisation. A high genetic diversity and an absence of evidence of inbreeding suggests the regional decline of Scots pine did not critically reduce allelic variation. The post-glacial relationship between Irish and Scottish pine is discussed, and a suggestion from recent palaeoecological work that reintroduced Scots pine be managed as a native species is now further supported by genetic data.

## Background

Scots pine (*Pinus sylvestris* L.) occupies an enormous longitudinal range of about 14,000 km and is the only pine which is native to Ireland and Britain [1, 2]. Its distribution extends from Spain and Ireland (7°W) to the eastern coast of Siberia (138°E), and from 37°N in Spain to over 70°N in Norway [3–5]. Across this range strong local adaptation and clinal variation is reflected in efficient purifying selection in genic regions, although linkage disequilibrium (LD) decays quickly and there is high variation at neutral loci [6, 7]. A high level of shared polymorphisms at these loci and a large effective population size is a function of the highly efficient gene

<sup>†</sup>Samuel Belton and Philippe Cubry contributed joint first authorship.

\*Correspondence:

Colin T. Kelleher  
colin.kelleher@opw.ie

<sup>1</sup> DBN Plant Molecular Laboratory, National Botanic Gardens of Ireland, Glasnevin, Dublin, Ireland

<sup>2</sup> DIADE, Univ de Montpellier, CIRAD, IRD, Montpellier F-34090, France

<sup>3</sup> National Parks and Wildlife Service, Department of Housing, Local Government and Heritage, 90 King Street North, Smithfield, Dublin, Ireland



© The Author(s) 2024. **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/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

flow displayed by this species [6]. Scots pine is a non-specialist, pioneer species which is most competitive in poorer, marginal soils [3, 8]. Economically, it is one of the most commercially important species in Europe [9]. In Ireland, it occupies c. 1.2% of total forest area [10].

The post-glacial history of Scots pine is complex because of its broad range and high adaptability [11–13]. Unlike most broadleaf species, Scots pine persisted outside of the Mediterranean peninsulas during the last glacial maximum (LGM) [14–16] and evidence exists which indicate that some conifers are likely to have survived in cryptic northern refugia adjacent to glacial fronts [17, 18]. However, the extent to which these may have contributed to the current distribution of Scots pine in northwest Europe is unclear and most evidence converges on a scenario in which it was recolonised after the LGM from a southern direction, most likely from the largest refugia in the Iberian Peninsula, Italy or the Balkans, initially forming pioneer forests with *Betula* spp. [16, 19, 20].

Scots pine arrived in southern England c. 11,470 cal BP (calibrated radiocarbon years before present, as given in the literature) and in northern Scotland by 4,470 cal BP [1, 5, 21]. A presence in southwest Ireland is recorded at 10,740 cal BP [5, 22], though its arrival is thought to have been from across the Celtic Sea rather than through Britain [21, 23]. In Ireland, Scots pine became dominant in western and upland sites in the early Holocene. In north-central Ireland, its occurrence was more limited due to competition from *Ulmus*, *Quercus* and *Corylus*, although isolated populations on lowland peatland sites persisted [24]. By 4,500 cal BP, it was mainly restricted to mountains and raised bogs, where it became further displaced by blanket peat formation and wetter hydrological conditions, respectively [1, 24, 25]. This was part of a wider contemporaneous decline of the species across northwest Europe [26–28].

Scots pine continued to persist in isolated pockets [29, 30], but was extirpated from multiple countries in northwest Europe during the late Holocene [20, 24, 31, 32]. This was assumed to have occurred in Ireland by c. 1,600 cal BP, but recent evidence from a putative native relict Scots pine population (Rockforest) in western Ireland has challenged this assumption [5, 33]. Scots pine survived in Scotland [34], and other relict populations may occur elsewhere in Britain [35, 36]. In Ireland, many non-relict populations occur which are the result of 19th and 20th century plantations from Scotland. However, some are naturalised and support vegetation communities with similarities to fossil pinewood assemblages from Ireland and extant native pinewoods elsewhere in northwest Europe [37]. Scots pine has been included in the Irish Native Woodland

Scheme [38] and it has been recommended that reintroduced populations be managed as a native species [37, 39].

The relatively extensive ecological and palaeological characterisation of Irish pinewoods means that a genetic analysis is now long overdue, as concluded by other works [33, 40]. This will help to inform conservation approaches, which are important as there is growing evidence illustrating a high vulnerability of the species to climate change, and extreme shifts in its distribution range are predicted [41–45]. In this paper, Irish Scots pine are genetically characterised for the first time. In doing so, the question of whether the regional decline in Scots pine has resulted in critically low levels of genetic diversity is addressed. Results also shine light on the genetic composition and origins of Scots pine in Ireland.

## Materials and methods

### Plant material and DNA extraction

Sixteen Irish populations detailed from previous ecological surveys were selected for sampling [8, 46], along with three additional northern populations to improve geographical coverage (Fig. 1). The origin and planting date of each, where known, are given in Table S1. Material from two French and one Spanish population were also sampled, as well as four Scottish origin and three Norwegian origin genotypes. Spanish samples were collected from a native population, whereas the French samples were collected from naturalised populations from an unknown origin. The four Scottish genotypes came from a breeding programme and neither them nor the Norwegian samples were collected from discrete sampling populations. DNA was extracted as described in Belton et al. [47].

### Mitochondrial polymorphism

High resolution melting analysis (HRM) [48] was used to study mitochondrial DNA (mtDNA) variation at the *nad7* intron1 and *nad1* B/C loci [15, 49]. The former was targeted using primers H and I from Soranzo et al. [49], and the latter was targeted with internal primers (Nad7\_indel\_F: GCTGCCCCATTCAATTACAC; Nad7\_indel\_R: ATAAGGAGAGGATGCGGAAG) designed from sequences published by Pyhäjärvi et al. [15]. PCRs were performed using a QIAGEN Type-it HRM PCR kit (cat. No. 206,544) with 1 µl (5 ng) DNA working solution, 1 µl DNase-free water and 3 µl master mix on a QIAGEN Rotor-Gene Q machine using a default HRM protocol. Fluorescence was monitored continuously with an endpoint melting range of 65–95 °C and alleles were called manually using the manufacturer's software.



**Fig. 1** Locations/provenances of Scots pine sampling populations used in the current study

### Chloroplast SSR analysis

Eight mononucleotide repeat chloroplast SSRs (cpSSRs) were selected from the literature [50–52]. These were combined into two multiplex reactions (Table S2). PCRs were prepared using a Type-it Microsatellite PCR kit (QIAGEN; cat. No. 206,243), with each comprising 3 µl Type-it master mix, 2.5 µl (25 ng) DNA working solution, 1 µl primer mix and 3.5 µl DNase-free water. PCR conditions were set according to the kit quick protocol. Allele

size was determined using the STRand software (UCDavis Veterinary Genetics Laboratory, Davis, USA) after electrophoresis on an ABI 3730 DNA analyser. The Scottish and Norwegian genotypes as well as one French population (Plan-d'Aups) were not included in this analysis.

### Nuclear SSR analysis

In total, 18 nuclear SSRs (nSSRs) were selected from the literature and arranged into three multiplexes [49,

51, 53–57] (Table S3). Each PCR contained 4.5 µl DNA working solution (22.5 ng), 1 µl primer mix, 7.5 µl Type-it master mix and 2 µl DNase-free water. PCRs and allele calling were the same as for the cpSSR analysis.

## Data analysis

### Genetic diversity and differentiation (cpSSR)

Haplotypes and the frequencies of their private occurrences were determined using a custom R (v.4.2.0) script (Additional File 1), using the R packages *tidyr* (v.1.3.0) [58] and *adeigenet* (v.2.1.5) [59, 60]. *Adegenet* was also used to handle all raw allelic data prior to statistical analyses. Genetic diversity was calculated according to Nei's [61] unbiased genetic diversity index ( $H_{CP}$ ) using the software Haplotype Analysis (v.1.05) [62]. Allelic richness was calculated by rarefaction to seven alleles using the R package *PopGenReport* (v.3.0.4) [63, 64]. Genetic distance between individuals was calculated according to Nei [65] and Bruvo et al. [66] using the R package *poppr* (v.2.9.3) [67, 68].

Differentiation estimates ( $F_{ST}$ ) were calculated according to Nei [69] and Weir and Cockerham [70] using the R package *hierfstat* (v.0.5-7) [71]. Estimates of statistical significance were derived by bootstrapping using the R package *mmod* (v.1.3.3) [72]. Confidence interval (CI) values were, in some cases, converted to  $p$  values according to Altman and Bland [73]. Variation components of the data were determined using an Analysis of Molecular Variance (AMOVA), with significance determined from 1000 permutations of the data using the *poppr* wrapper of the AMOVA function in the R package *ade4* (v.1.7–18) [74].

### Genetic diversity and differentiation (nSSR)

Tests for LD were performed by determining the pairwise index of association ( $\bar{r}_d$ ) [75] for each of 999 permutations of the data using *poppr*. Tests for departure from Hardy-Weinberg equilibrium (HWE) were performed using the exact test of HW proportions based on 1000 Monte Carlo permutations of alleles [76], as implemented in the R package *pegas* (v.1.0–1) [77]. As some nSSR loci were markers transferred from other species, a high frequency of null alleles was expected. Frequencies were estimated according to the expectation maximisation algorithm [78] implemented in the R package version (v.1.1.7) of *GENEPOP* [79]. Population genetic diversity was inferred by calculating expected heterozygosity ( $H_e$ ) using *adeigenet*. Private alleles were identified using *poppr*, and allelic richness calculated by rarefaction to 14 nSSR alleles.

$F_{ST}$  analysis was computed as before, whereas differentiation was also estimated based on allele size variance ( $R_{ST}$ ) according to Slatkin [80] using *pegas*. Inbreeding

coefficients ( $F_{IS}$ ) were calculated according to Nei [69] and Weir and Cockerham [70]. The effect of null alleles on  $F_{IS}$  was estimated using a Bayesian approach implemented in the programme *INEST* (v.2.0) [81]. Specifically, a null allele model was compared against a model which also includes  $F_{IS}$  to describe the data. This procedure is outlined in detail in Belton et al. [82].

Genetic distance between individuals was again calculated using Nei's and Bruvo's genetic distance using *poppr*. As the latter assumes a stepwise mutation model (SMM), some loci were removed because they did not fulfil this assumption. Using the programme *ML-RELATE* [83], relatedness ( $r$ ) was also estimated using a maximum likelihood (ML) procedure which accommodates null alleles [84]. In practice,  $r$  was estimated after readjusting allele frequencies according to the presence of null alleles, after which  $r$  was estimated under ML. *ML-RELATE* was also used to calculate the likelihood of four categories of pedigree relationship (unrelated, half-sibling, full-sibling and parent - offspring).

### Analysis of population structure (cpSSR and nSSR)

The main genetic and spatial structure of the cpSSR haplotypes was investigated by performing a Spatial Analysis of Molecular Variance (SAMOVA) using *SAMOVA* (v.2.0) [85]. This approach, which attempts to maximise differentiation between groups of populations, was performed with and without geographic data. Simulations were conducted using default parameters for SSR data, with  $K$  ranging from two to 10 and with the annealing process for each repeated 100 times. Sub-population structure based on nSSR variation was analysed using *STRUCTURE* (v.2.3.4) [86], which was run across 20 CPU cores via the programme *StrAuto* (v.1.0) [87]. The admixture model with correlated allele frequencies was tested using sample location information (LOCPRIOR). The MCMC simulation involved a burn-in of 100,000 and 500,000 retained iterations for each of  $K=15$  inferred clusters, with 20 replications for  $K$ . As *STRUCTURE* is sensitive to deviations from HWE, simulations were repeated after adjusting for null alleles (removal of loci displaying null allele frequencies  $\geq 10\%$ ). The optimal value for  $K$  was inferred using the  $\Delta K$  method [88] and replications were merged using the Greedy algorithm in the programme *CLUMPP* (v.1.1.2) [89]. The R implementation of *TESS3* (*tess3r* v1.1.0) was also used to infer population clusters with the use of spatial information. *TESS3* is a non-model-based approach which uses matrix factorisation instead of Bayesian inference and is not sensitive to LD or deviations from HWE [90].  $K=1$  to  $K=20$  populations were tested, each replicated 40 times. The optimal value for  $K$  was then determined by selecting the value with the lowest cross-validation score,



as recommended by the programme authors. Finally, in addition to pairwise  $F_{ST}$  and  $R_{ST}$  comparisons, population sub-division was further investigated by performing a Principal Coordinates Analysis (PCoA) based on Nei's genetic distance using adegenet.

Results

Mitochondrial variation

All samples exhibited the same *nad7* intron 1 allele. Only Spanish trees contained a different allele from other provenances at the *nad1* B/C intron. Samples from Ireland, Scotland, France and Norway were therefore all fixed for the most common European mitotype, which is mitotype *a*, whereas the Spanish samples analysed were all fixed for mitotype *c* (data not shown).

Chloroplast variation

Between two and seven alleles were identified per cpSSR locus (Table 1). The distributions of allele size variants were unimodal for each, with minor alleles belonging to the extremes of the allele size ranges. Variation at the cpSSR level was very high; a total of 134 haplotypes were detected, 122 of which occurred in Ireland (Table S1). The highest frequency (6.61%) for any one haplotype was H15. An average of 3.76 private haplotypes were detected, with frequencies being highest for Cavan and lowest for Vale of Clara and Knockastackeen, which harboured none (Fig. 2). Allelic richness was highest in Ballykelly and lowest in Dale Wood, which also had the lowest  $H_{CP}$  (Fig. 2). Overall  $H_{CP}$  was 0.965 ( $\pm 0.063$  S.D.), or 0.967 ( $\pm 0.065$  S.D.) when removing continental populations. Altogether, the number of different alleles and genetic diversity was lowest in Dale Wood, whereas no single population stood out for having a particularly high level of either.

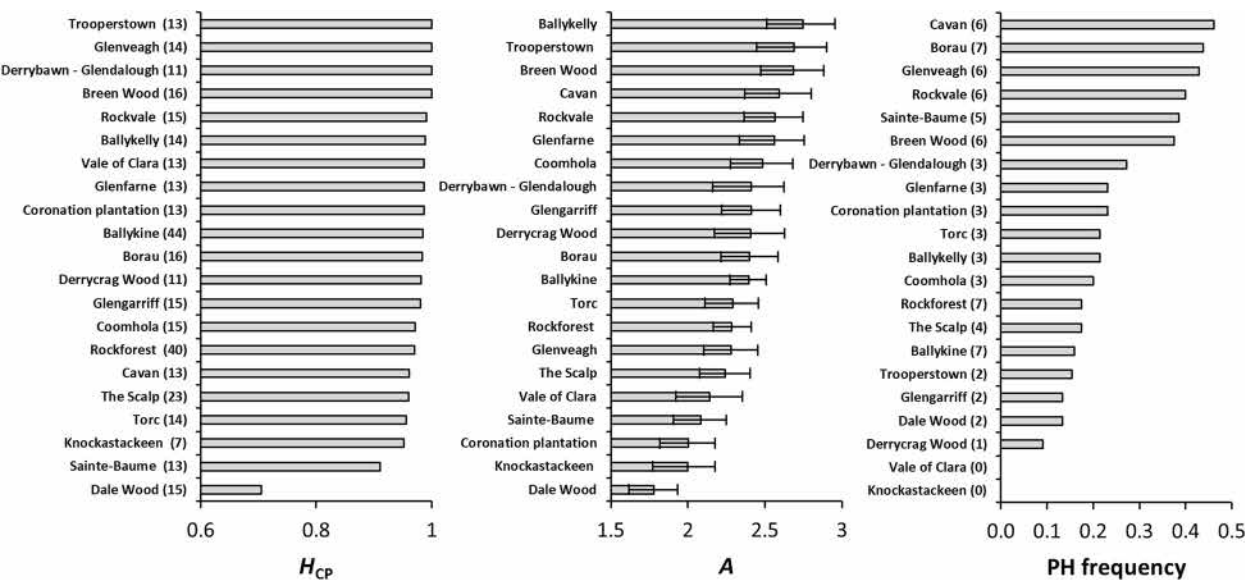
Mean Nei genetic distance between individuals calculated at the population level ranged from 0.359 (Dale Wood) to 0.811 (Rockvale), or from 0.176 (Dale Wood) to 0.300 (Ballykelly) when using Bruvo's genetic distance (Figure S1). As the latter assumes a SMM, it may be a more realistic description of haplotype relatedness compared to Nei's standard genetic distance, which assumes only an Infinite Alleles Model. Overall  $F_{ST}$  was very small but significantly greater than zero ( $F_{ST} = 0.020$ , 95% CI: 0.001–0.041, or  $p = 0.05$ ; Table 1). This significance was mainly attributable to Pt30204, which was the only cpSSR locus which displayed an  $F_{ST}$  value which was significantly greater than zero (Table 1). Only a sample size-independent estimate ( $F_{ST} = 0.019$ , 95% CI: 0.001–0.037, or  $p = 0.046$ ) was significant when just Irish populations were considered. Pairwise  $F_{ST}$  comparisons showed that this was almost entirely due to differentiation between Dale Wood and all other populations (Figure S2). AMOVA showed that only 2.47% of total variation was between populations (or 2.41% when only Irish populations were considered; Table 2).

Nuclear variation

Six nSSR loci had null allele frequencies exceeding 10% (Table 3) and subsequent analyses were performed before and after removing them. All loci were considered unlinked, as no significant LD ( $p \leq 0.05$ ) was found for 3,672 pairwise comparisons after correcting for familywise error (Holm and Bonferroni). Following Bonferroni correction, significant departure from HWE was detected for 20 out of 432 cases, corresponding to HWE departures in nine populations (Table 4). This dropped to two out of 288 tests after adjusting for null alleles. Null alleles also inflated initial  $F_{IS}$  estimates (Table 4), and only Coomhola displayed a significant (albeit low)

**Table 1** Diversity and differentiation statistics of the eight cpSSR markers used to genotype Scots pine (*Pinus sylvestris*) individuals ( $n = 348$ ). Statistics include number of alleles (N), allelic richness (A), intra-population diversity ( $h_s$ ), total diversity ( $h_T$ ), diversity which apportioned between populations ( $F_{ST}$ ), with (Nei) and without adjusting for sample size (WC).  $F_{ST}$  values which are marked with an asterisk are significantly different from zero ( $p \leq 0.05$ )

	N	Allele size range	A	$h_s$	$h_T$	$F_{ST}(\text{Nei})$	$F_{ST}(\text{WC})$
Pt1254	5	59–63	1.400	0.384	0.388	0.011	0.012
Pt41093	2	73–74	1.118	0.124	0.125	0.013	0.008
Pt26081	3	107–109	1.385	0.360	0.363	0.008	0.016
Pt15169	7	123–129	1.709	0.707	0.726	0.027	0.026
Pt30204	6	137–142	1.670	0.671	0.712	<b>0.058*</b>	<b>0.051**</b>
Pt71936	4	142–145	1.281	0.275	0.281	0.021	0.016
Pt36480	6	142–147	1.618	0.624	0.640	0.026	0.027
Pt87268	5	161–165	1.257	0.273	0.273	0.000	0.000
mean	4.75	-	1.430	0.427	0.439	<b>0.020*</b>	<b>0.019*</b>



**Fig. 2** Chloroplast genetic diversity ( $H_{CP}$ ), allelic richness ( $A$ ) and the frequency of private haplotypes (PH) for each Scots pine population (excluding Norwegian and Scottish samples).  $A$  was calculated based on rarefaction down to seven cpSSR alleles. Values in brackets in the  $H_{CP}$  and PH frequency plots are sample number and the number of detected PHs, respectively. Error bars on  $A$  values are 95% CIs derived from 1000 bootstrap permutations of the data

**Table 2** Results of an AMOVA based on cpSSR variation. AMOVA results are given based on all sampled trees (All) and based on trees sampled in Ireland only

Components (All)	d.f	SSD	Variance	Variance (%)	p value
Between popula-tions	20	97.146	0.087	2.472	0.001
Within populations	327	1123.182	3.435	97.528	< 0.001
Total	347	1220.328	3.522	100.000	
Components (Ireland)	d.f	SSD	Variance	Variance (%)	p value
Between popula-tions	18	87.461	0.085	2.406	0.002
Within populations	300	1035.893	3.453	97.594	< 0.001
Total	318	1123.354	3.538	100.000	

null allele-adjusted  $F_{IS}$  value. Overall,  $F_{IS}$  for Ireland was very low, at 0.022 (95% highest density interval: 0.0049–0.0397).

Diversity ( $H_e$ ) and allelic richness ( $A$ ) were highest for Derrybawn – Glendalough and Borau, and lowest for Coronation Plantation and Dale Wood respectively (Fig. 3). These differences remained after adjusting for null alleles (Figure S3). For all Irish populations,  $H_e$  was 0.559 (or 0.540 after adjusting for null alleles).

Differentiation at the nSSR level ( $F_{ST} = 0.032$ , 95% CI: 0.022–0.037, or  $p < 0.001$ ; Table 3) was not significantly greater than that observed for the cpSSR loci. When removing non-Irish populations,  $F_{ST}$  remained

low but significant ( $F_{ST} = 0.018$ , 95% CI: 0.014–0.023, or  $p < 0.001$ ). Null alleles did not affect these estimates. Pair-wise comparisons revealed that differentiation was highest between Plan-d'Aups and Dale Wood ( $F_{ST} = 0.149$ , 95% CI: 0.120–0.181, or  $p < 0.001$ ; Fig. 4). Although private allele frequency was generally highest for continental populations (Table 4), differentiation based on allele size (rather than allele frequency) was overall not significant ( $R_{ST} = 0.011$ , 95% CI: 0.000–0.065, or  $p = 0.518$ ; Figure S4). AMOVA also showed that a similarly low proportion (3.84%) of variation was due to between-population differences, although the contribution of non-Irish populations appeared to be greater as removing them reduced the between-population variance to 2.73% (Table 5).

Mean Nei genetic distance between individuals ranged from 0.463 (Coronation Plantation) to 0.670 (Torc; Figure S5a). As for cpSSR genetic distance, mean nSSR genetic distance based on allele size differences (*i.e.*, assumption of a SMM) yielded different estimates; after removing SPAG7\_14 and psy144, which violated the SMM, mean Bruvo's genetic distance based on nSSR alleles ranged from 0.338 (Plan-d'Aups) to 0.421 (Scottish (Coillte); Figure S5b). Out of 58,996 pairwise comparisons, the distribution of ML estimates for  $r$  reduced rapidly, with 83.26% being 0.1 or less (Figure S6). Distributions were less skewed for Dale Wood, Plan-d'Aups and Sainte-Baume (Figure S7), populations which also showed greater mean values of  $r$  per tree (Fig. 5). Dale Wood and Plan-d'Aups also had the highest frequency of full and half siblings,

**Table 3** Summary statistics for the 18 nSSR loci used to genotype Scots pine. Abbreviated values from left to right include the total detected alleles (N), allelic richness (A), observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ), inbreeding coefficient ( $F_{IS}$ ) and genetic diversity which apports between populations ( $F_{ST}$ ).  $F_{IS}$  and  $F_{ST}$  values are calculated according to Nei (1987) and Weir & Cockerham (1984). The latter are unbiased by sample size differences.  $F_{IS}$  and  $F_{ST}$  values in bold are significantly greater than zero. (\* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ ). Loci with null frequency values in bold were removed from analyses where indicated

Locus	N	Allele size range	A	$H_o$	$H_e$	$F_{IS}$ (Nei)	$F_{IS}$ (WC)	$F_{ST}$ (Nei)	$F_{ST}$ (WC)	Null frequency
PtTX3116	26	116–278	1.803	0.458	0.869	<b>0.453***</b>	<b>0.459***</b>	<b>0.038**</b>	<b>0.030***</b>	<b>0.175</b>
PtTX3032	49	275–512	1.916	0.874	0.954	<b>0.088***</b>	<b>0.076***</b>	0.012	<b>0.010*</b>	0.026
psy119	5	310–322	1.130	0.100	0.104	0.013	0.019	0.035	0.027	0.005
SPAC11.6	45	108–220	1.881	0.615	0.926	<b>0.346***</b>	<b>0.328***</b>	0.016	<b>0.016**</b>	<b>0.135</b>
psy117	8	215–229	1.642	0.230	0.691	<b>0.672***</b>	<b>0.659***</b>	<b>0.030*</b>	<b>0.030**</b>	<b>0.251</b>
SPAC12.5	30	122–190	1.879	0.901	0.914	-0.016	-0.010	<b>0.020*</b>	<b>0.021***</b>	0.019
SPAG7.14	55	174–244	1.900	0.808	0.944	<b>0.117***</b>	<b>0.126***</b>	<b>0.025***</b>	<b>0.024***</b>	0.037
PtTX3107	12	144–171	1.723	0.379	0.765	<b>0.465***</b>	<b>0.503***</b>	0.009	0.004	<b>0.182</b>
PtTX4001	15	196–226	1.755	0.762	0.781	0.014	0.005	<b>0.027*</b>	<b>0.023**</b>	0.021
PtTX4011	9	231–277	1.530	0.441	0.583	<b>0.227***</b>	<b>0.207***</b>	<b>0.062***</b>	<b>0.050***</b>	0.075
psy118	7	287–317	1.134	0.111	0.120	0.058	0.068	0.009	0.009	0.010
psy12	4	197–206	1.303	0.255	0.302	<b>0.140*</b>	<b>0.141*</b>	0.024	0.019	<b>0.769</b>
psy125	2	212–215	1.042	0.038	0.043	0.087	0.086	0.019	0.034	<b>0.312</b>
psy136	5	244–256	1.098	0.122	0.117	-0.038	-0.045	0.000	0.000	0.000
psy142	6	165–175	1.648	0.659	0.677	-0.002	0.011	0.020	0.018	0.030
psy144	5	164–175	1.148	0.120	0.144	0.043	0.078	<b>0.103*</b>	<b>0.105**</b>	0.014
psy157	6	184–199	1.388	0.369	0.414	0.036	0.050	<b>0.086***</b>	<b>0.066***</b>	0.021
SPAC11.4	20	132–170	1.811	0.842	0.852	-0.015	-0.015	<b>0.035***</b>	<b>0.025***</b>	0.016
Mean	17	-	1.541	0.449	0.567	<b>0.149***</b>	<b>0.156***</b>	<b>0.032***</b>	<b>0.028***</b>	0.117

and the lowest frequency of unrelated individuals (Figure S8).

### Population structure

SAMOVA based on cpSSR variation revealed that the main structure was between Dale Wood and all other populations (*i.e.*,  $K=2$ ;  $F_{CT} = 0.0828$ ,  $p=0.0489$ ), among which differentiation was not significant ( $F_{SC} = 0.0141$ ,  $p=0.0860$ ; Table S4a). Glenfarne separated as a third group at  $K=3$  and Knockastackeen as a fourth at  $K=4$ . At the latter, Torc also clustered with Dale Wood, whereas  $F_{SC}$  was negative at  $K \geq 5$ . When the data were geographically constrained, results remained the same until  $K \geq 9$ , after which group assignment differed (Table S4b). PCoA also showed that Dale Wood clearly separated along the main axis, which explained 59% of the variation (Figure S9), which is in agreement with pairwise  $F_{ST}$  and SAMOVA results. No approach showed any clear distinction between Irish and continental populations.

By contrast, STRUCTURE analysis of the nSSR data showed that the Irish samples and the four Scottish genotypes were clearly differentiated from the continental populations (including the three Norwegian genotypes) at  $K=2$  and  $K=3$  (Fig. 6). Derrybawn – Glendalough displayed a relatively high level of admixture with a gene

pool that was most frequent on the continent, whereas Torc and especially Dale Wood in the southwest of Ireland displayed a disproportionate frequency of a third ancestral group at  $K=3$ . Both axes of a PCoA plot separated the Irish and continental populations, suggesting that the main structure is described at  $K=2$  (Fig. 7). As STRUCTURE is sensitive to deviations from HWE, the analysis was repeated after adjusting for null alleles, with the optimal value for  $K$  indeed being inferred at  $K=2$  ancestral groups (Fig. 8, Figure S10). This was also more apparent based on a null allele-adjusted PCoA (Figure S11). At this level of structure, admixture with continental trees was inferred to be lower. Continental and several Derrybawn – Glendalough trees also separated as before, with Torc and Dale Wood becoming differentiated again at  $K=3$ . Interestingly,  $\Delta K$  was second highest at  $K=6$  (Fig. 8b), for which one genotypic group was more frequent in Rockforest trees compared to all other Irish trees (Fig. 8c). This was the case for each of  $K=5$  to  $K=15$  genotypic groups (Figure S12), indicating a very small shift in allele frequency in Rockforest trees.

TESS3, which is not sensitive to null alleles but uses spatial coordinates to assist clustering, also suggested  $K=2$  as the main structure (Figure S13). Unlike with STRUCTURE however, this separated all Irish trees such

**Table 4** Geographic coordinates for Scots pine sampling populations. Also indicated is the number of trees sampled per population (N), the frequency of private alleles (Pa) detected, Nei's inbreeding coefficient ( $F_{IS}$ ) and null allele-adjusted  $F_{IS}$  ( $F_{IS}INEST$ ). Values which are significantly different from zero are marked with asterisks (\* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ ).  $F_{IS}$  values in bold indicate significant departure from HWE and  $F_{IS}INEST$  values in bold indicate that an inbreeding model was favoured over a null allele model

Population	Longitude	Latitude	N	Pa	$F_{IS}$	$F_{IS}INEST$
Ballykelly	-7.0487	55.0434	13	0.0000	0.2121**	<b>0.0824</b>
Ballykine	-9.3461	53.5585	43	0.0078	<b>0.2025***</b>	<b>0.0465</b>
Borau	-0.5882	42.6581	12	0.0278	0.1831**	<b>0.0387</b>
Breen Wood	-6.2376	55.1364	13	0.0256	0.2024**	<b>0.0418</b>
Cavan	-7.2172	53.9289	13	0.0085	0.1733*	<b>0.0273</b>
Coomhola	-9.4662	51.7759	13	0.0085	<b>0.1959**</b>	<b>0.0781*</b>
Coronation plantation	-6.3598	53.1544	14	0.0040	<b>0.1905**</b>	<b>0.0589</b>
Dale Wood	-9.6277	51.9678	13	0.0043	0.2241*	<b>0.0293</b>
Derrybawn - Glendalough	-6.3268	53.0065	16	0.0208	0.2395***	0.0707
Derrycrag Wood	-8.3941	53.0420	9	0.0000	0.0979	0.0399
Glenfarne	-7.9691	54.3065	14	0.0119	<b>0.2099*</b>	0.0203
Glengarriff	-9.5807	51.7538	15	0.0000	<b>0.1956*</b>	0.0482
Glenveagh	-7.9437	55.0557	14	0.0000	0.1332*	<b>0.0430</b>
Knockastackeen	-8.0917	52.3968	7	0.0000	0.1454*	0.0403
Norway <sup>a</sup>	8.5	60.4	3	0.0370	0.2381*	0.0345
Plan-d'Aups	5.7170	43.3300	7	0.0238	0.2067*	<b>0.0700</b>
Rockforest	-8.9726	53.0022	35	0.0063	<b>0.1622*</b>	0.0191
Rockvale	-8.9106	53.0303	15	0.0185	0.1851*	0.0122
Sainte-Baume	5.8640	43.4525	8	0.0347	0.2551**	0.0862
Scottish (Coillte) <sup>a</sup>	-4.2	56.4	4	0.0000	0.1741*	0.0429
The Scalp	-6.1786	53.2176	23	0.0072	<b>0.1784**</b>	0.0409
Torc	-9.5079	51.9964	14	0.0040	<b>0.2519***</b>	0.0719
Trooperstown	-6.2736	53.0090	12	0.0093	0.1023*	<b>0.0489</b>
Vale of Clara	-6.2562	52.9674	14	0.0040	<b>0.2201***</b>	<b>0.1089</b>

<sup>a</sup> These are groups of individuals which originate from the same country but not necessarily the same forest site. Geographic coordinates for these are country level only and are given for GIS plotting

that neither Derrybawn – Glendalough, Torc nor Dale Wood were differentiated at either  $K=2$  or  $K=3$ . TESS3 also did not infer any additional substructure at  $K \geq 4$  (data not shown).

## Discussion

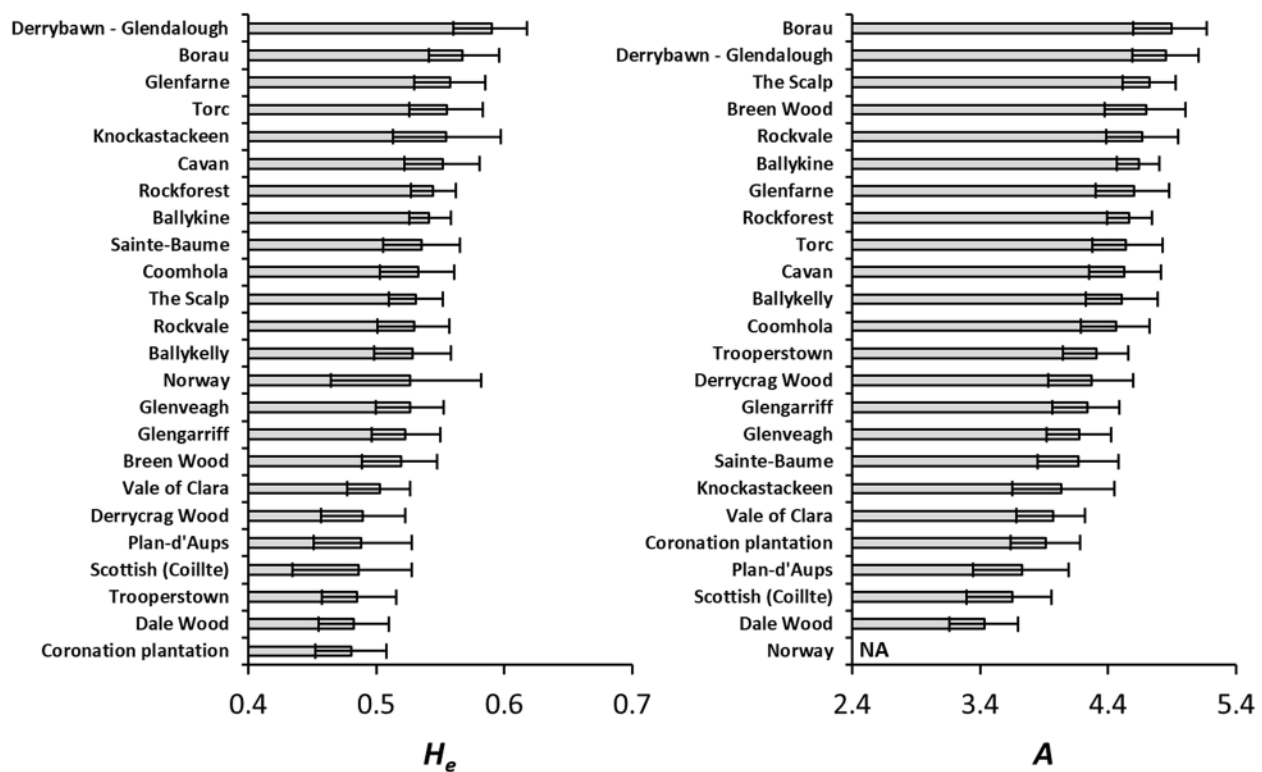
### Genetic diversity of Irish Scots pine in a European context

The absence of any diversity at the selected mtDNA loci is in line with other works, which show that these alleles are fixed across most of the Scots pine distribution range [14, 91]. Unlike mtDNA, conifer chloroplast DNA (cpDNA) has an intrinsically higher substitution rate and is dispersed via pollen, which gives rise to much greater variation [92–95]. The variation observed here at cpSSR loci was high (Table 1). Many of these loci have been used in other works to produce diversity estimates for populations of Scots pine across its distribution range (Table S5, Table S6).  $H_{CP}$  for Irish Scots pine ( $H_{CP} = 0.967 \pm 0.065$  S.D.) was similar to a range-wide average ( $H_{CP} = 0.888 \pm 0.109$  S.D.; Table S5).

This is probably a relatively reliable comparison, as re-estimations of  $H_{CP}$  for Irish trees using different subsets of cpSSR loci in accordance with different works are comparable with estimates for autochthonous pine in most regions across the distribution range (Table S6). Compared to some regions – Lithuania, Crimea, Caucasus and Asia Minor –  $H_{CP}$  for Ireland is even higher (Table S6).

The choice of nSSR loci has been much less standardised across works. Here,  $H_e$  estimates for each of 18 nSSR loci varied between 0.043 and 0.954 (Table 3), and produced an overall null allele-adjusted value of 0.540 (95% CI: 0.532–0.548) for Irish Scots pine. By contrast, Scalfi et al. [51] used just three highly variable loci (SPAC12.5, SPAG7.14 and SPAC11.4;  $H_e = 0.63$ –0.93) to estimate a value of 0.81 for Scots pine populations in northern Italy.  $H_e$  for Irish populations using this subset was 0.894 (95% CI: 0.886–0.901), and in this work, Spanish and French Scots pine did not differ from Irish trees when using the full loci set (Figure S14).





**Fig. 3** Genetic diversity ( $H_e$ ) and allelic richness ( $A$ ) based on nSSR variation in Scots pine sampling populations.  $A$  was calculated based on rarefaction down to 14 alleles. Error bars are 95% CIs derived from 1000 bootstrap permutations of the data

In general therefore, it appears that even though most Irish Scots pine come from reintroduced plantations, levels of genetic diversity are at least comparable – if not greater – than native pinewoods elsewhere. This may be partly due to plantations being sourced from Scotland, where diversity is particularly high ( $H_{CP} = 0.982-1$ ) [96, 97], potentially as a result of admixture between two refugial gene pools [98].

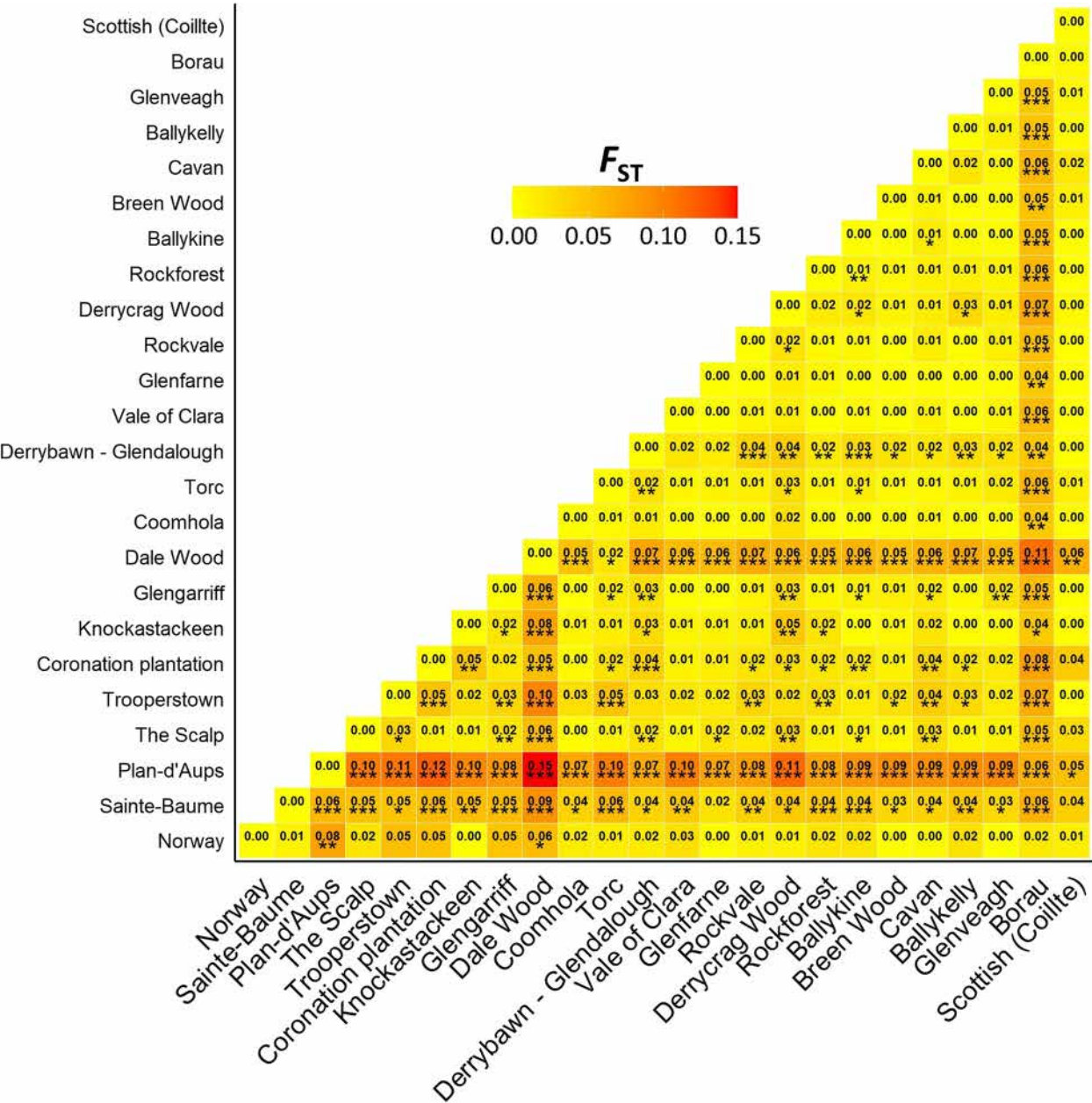
#### What is the genetic composition of Irish Scots pine?

A site in western Ireland (Rockforest, County Clare) was recently identified as a microrefugium of Scots pine, thus re-defining the natural range for the species [5, 33]. Palynological evidence indicates a continuous Scots pine presence at this site since at least 1,600 cal BP [5, 33], with macrofossil evidence pointing to a presence by at least 5,800 cal BP [39]. Pollen evidence from two locations within 40 km of Rockforest also suggests that parts of the west were colonised by *c.* 10,500 cal BP [22, 99]. As such, it appears that the Rockforest stand is a relict population of Scots pine, which previously occupied suitable sites across Ireland [21, 46]. It is therefore a good population to use for comparison with other populations of reintroduced Scots pine.

#### Insights from mtDNA loci

Scots pine's mitochondrial genome, which tracks seed dispersal, can be divided into five mitotypes (*a-e*) across its entire range and can be used to infer post-LGM and phylogeographic patterns [14, 91]. Excluding the limited number of Spanish trees which were fixed for mitotype *c*, all trees in the current study were fixed for mitotype *a*. This mitotype is also found throughout Spain, and is the most common mitotype across the species range, dominating northwest and central Europe, the Balkans, western Scandinavia and most of Siberia [14, 100, 101]. Mitotype *c* is generally restricted to northeast Spain and did not expand into northern Europe following the LGM [14, 41].

Improved phylogeographic resolution is now being provided by the discovery of additional mtDNA loci. For example, Wachowiak et al. [102] showed that Scots pine in its northern range contains at least eight mitotypes. Others have elucidated the composition of the Caucasus, and revealed that mitotype *a* in the Ural region can be divided into two subtypes [101, 103, 104]. Sinclair et al. [105] used the *coxI* gene as a probe to identify RFLPs in mtDNA from native Scottish pine. They identified three mitotypes, and a western-specific mitotype was hypothesised to be evidence of post-glacial migration from



**Fig. 4** Pairwise  $F_{ST}$  comparisons between sampled populations. Estimates are based on variation of nSSR allelic frequency according to Nei (1987). Values which are significantly different from zero are marked with asterisks (\* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ ;  $n = 344$ )

Ireland. However, the sequence variation defining these RFLPs currently remains unknown.

**Insights from cpSSR loci**

Although variation at cpSSR loci was relatively high, hardly any of this was between populations (Table 2, Table S4). This is also the case between Scottish and continental populations, and it is only between European and Anatolian populations that cpSSR structure is apparent

[96]. A recent establishment of northwest populations together with homogenising pollen-mediated gene flow likely explain the absence of sub-population differentiation in Europe [6, 93, 106, 107]. In this study, only Dale Wood was differentiated (Table S4, Figure S2), and this was due to very low variation (Fig. 2). This population (planted in 1829) was the most fixed for any one haplotype, presumably because it was derived from a limited seed source. Interestingly, the nearby Torc population

**Table 5** Results of a AMOVA based on nSSR variation. AMOVA results are given based on all sampled trees (All) and based on trees sampled in Ireland only

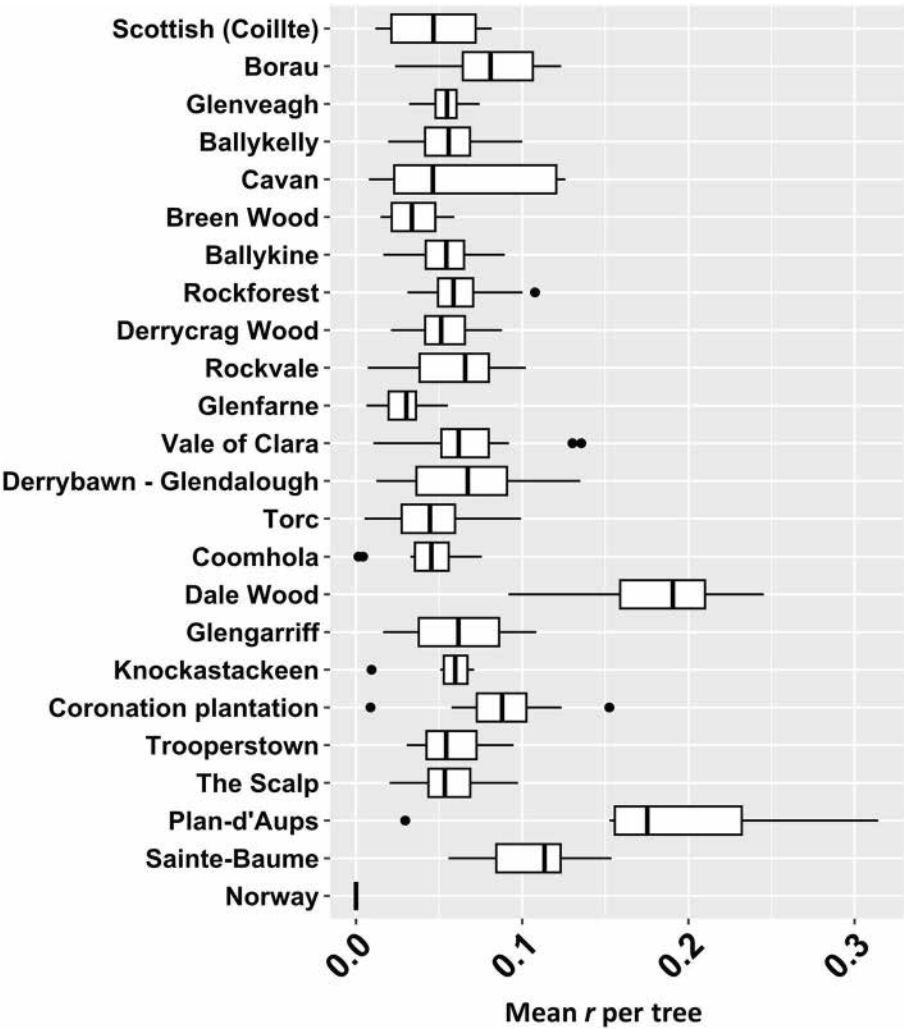
Components (All)	d.f	SSD	Variance	Variance (%)	p value
Between popula- tions	23	192.973	0.214	3.837	0.001
Within populations	319	1714.158	5.374	96.163	< 0.001
Total	342	1907.131	5.588	100.000	
Components (Ireland)	d.f	SSD	Variance	Variance (%)	p value
Between popula- tions	18	139.169	0.150	2.733	0.001
Within populations	290	1546.136	5.332	97.267	< 0.001
Total	308	1685.305	5.481	100.000	

(planted in 1880) was grouped with Dale Wood based on allelic composition at  $K=4$  cpSSR groups (Table S4).

Insights from nSSR loci

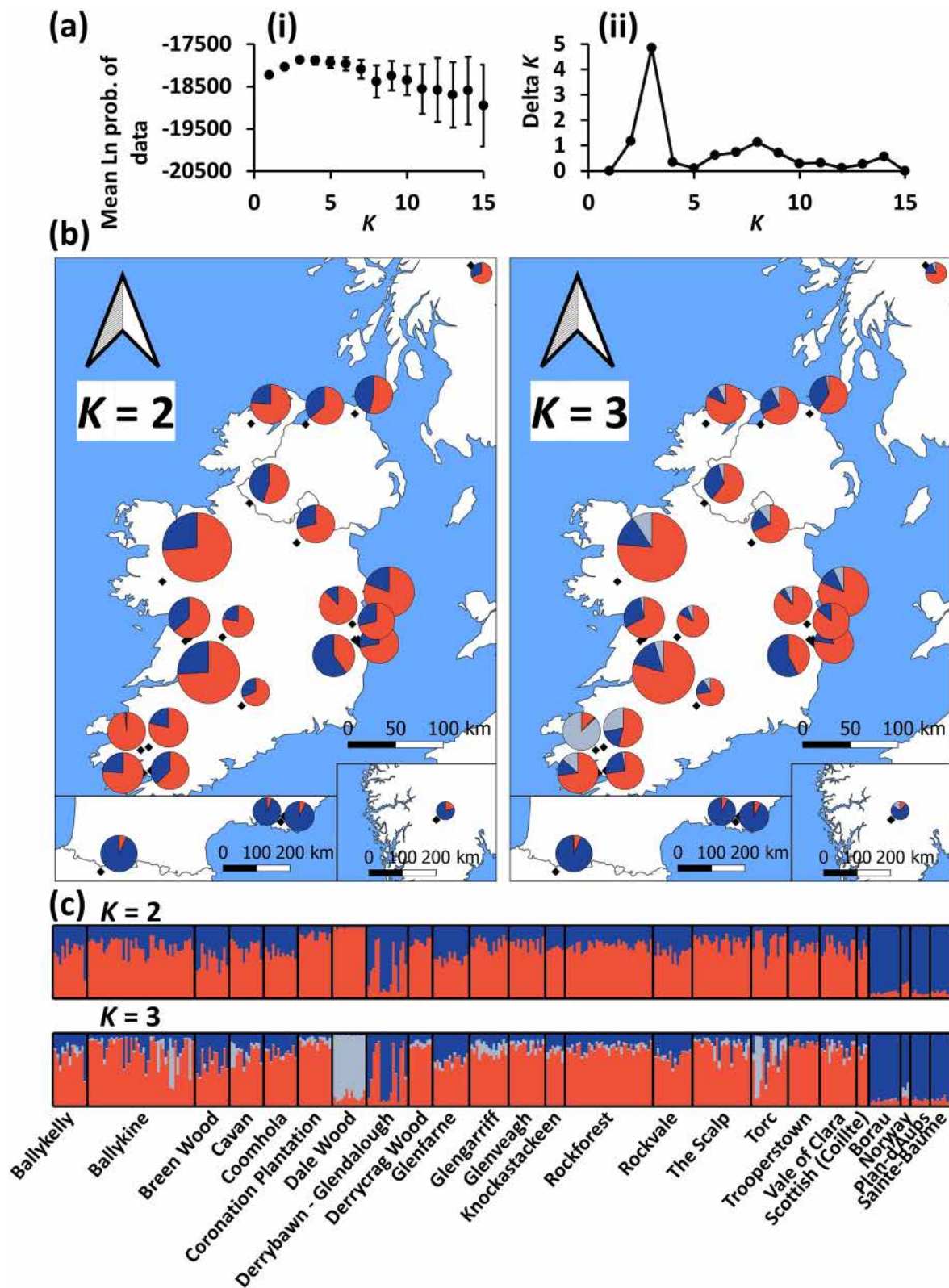
The between-population variation component at nSSR loci was slightly greater than at cpSSR loci (Table 5). This was due to differentiation between Irish and continental populations, with each belonging to a distinct gene pool (Fig. 6, Figure S10, Figure S13). Populations in Scotland are also differentiated from European populations at nDNA loci [98, 102], indicating that the Celtic Sea and English Channel act as relatively strong barriers to gene flow. Shifts in allele frequency giving rise to differentiation at cpSSR loci may occur more slowly due to lower variation and an absence of recombination.

Although 30 private alleles were detected in continental populations, the island and continental gene pools were characterised by variation in allele frequency ( $F_{ST}$ ; Fig. 4)



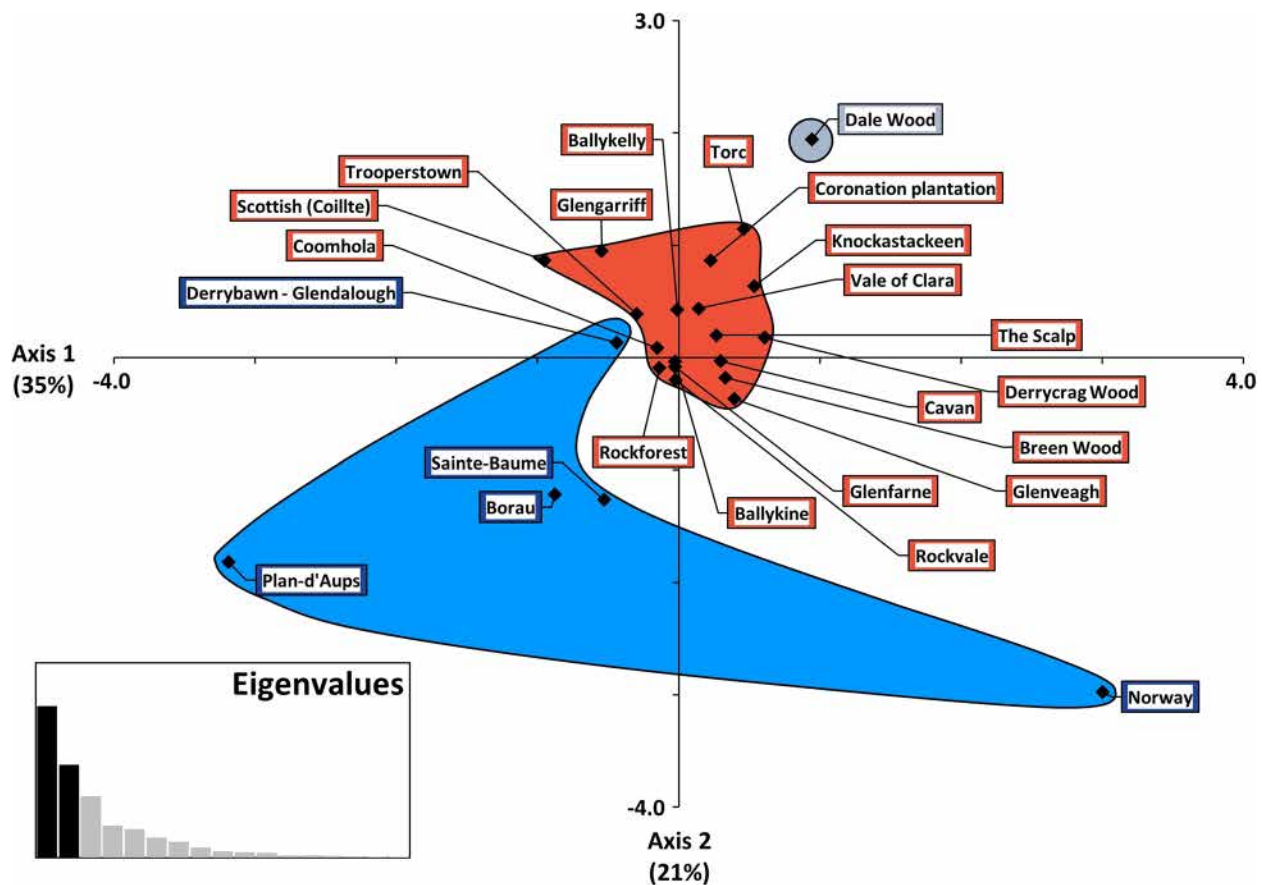
**Fig. 5** Boxplot showing mean values of  $r$  estimated under ML based on nSSR variation





**Fig. 6** Results of STRUCTURE analysis of Scots pine nSSR data. **a** (i) Mean posterior probability values of  $K$  ( $L(K)$ ), error bars are  $\pm$  S.D. (ii) Delta  $K$  values used to infer the main structure of the data. **b** Frequency of  $K$  groups within each sampling population, including those located on the continent (inset). Pie charts in Scotland and Norway do not indicate an exact sampling location, but are generic origins plotted for comparison's sake. **c** Bar plots of admixture coefficients for each sampled individual





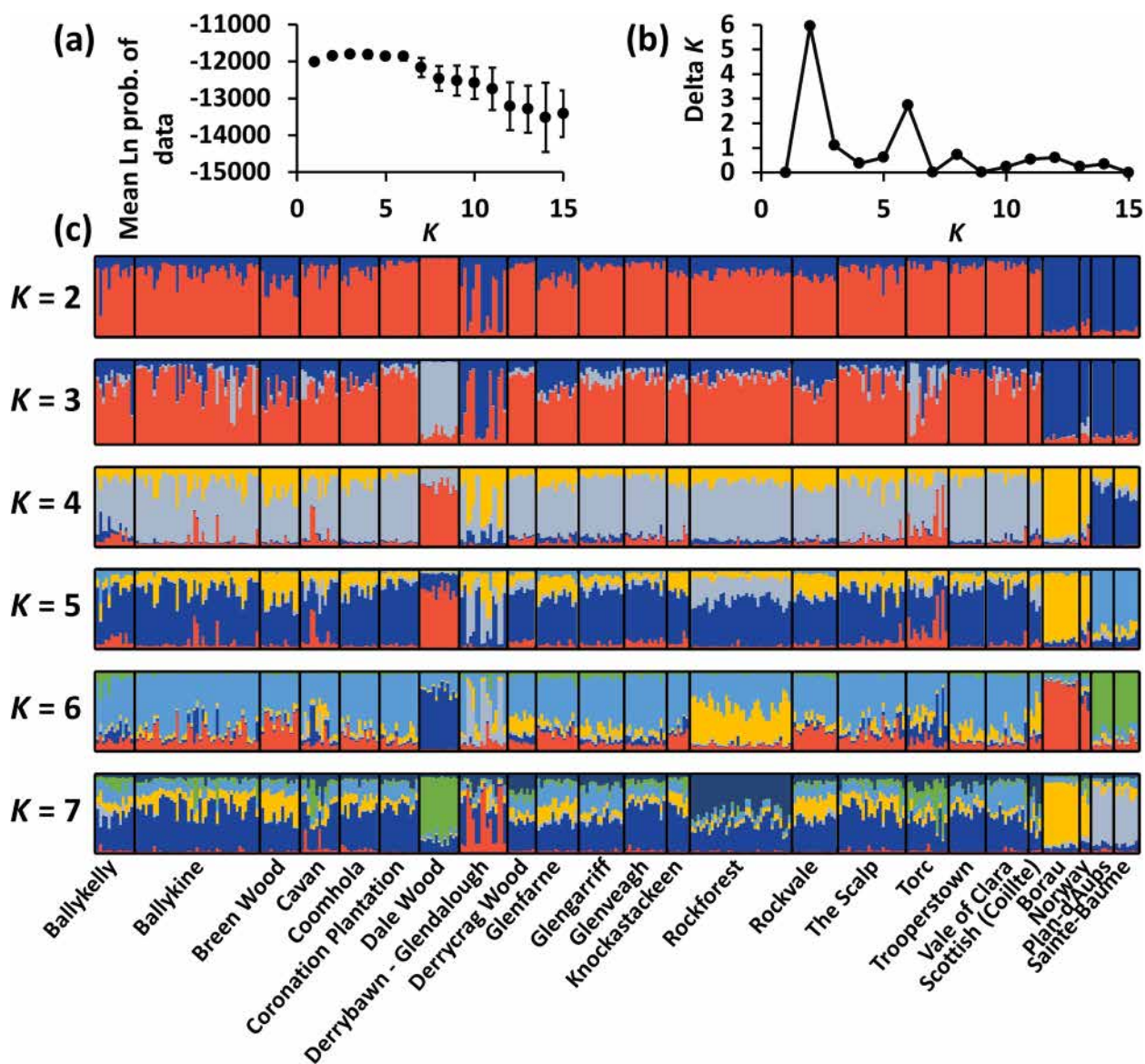
**Fig. 7** Principal Coordinates Analysis (PCoA) of the nSSR allelic composition of each Scots pine population. Shown are the first and second principal coordinates. Labels are coloured according to the STRUcTURE group ( $K=3$ ) that is most frequent in each population. Coloured shapes are also drawn around these points to assist in interpretation

rather than allele size ( $R_{ST}$ ; Figure S4). This may be partly shaped by the marginal differences in within-population relatedness observed between Irish and continental populations (Fig. 5). However, significant deviations from HWE were not detected in the continental populations whereas inbreeding was low in all populations (Table 4). Moreover, the two most differentiated populations (Dale Wood and Plan-d'Aups; Fig. 4) also harboured the highest frequency of related trees (Fig. 5). Therefore, these gene pools cannot only be a function of differences in within-population relatedness but are at least also reflective of a different demographic history.

Interestingly, geographic variability in allele frequency distributions among Scottish Scots pine has been suggested to signal historic admixture between two refugial gene pools during post-glacial recolonisation [98]. That admixture between the observed  $K=2$  gene pools here was greater in Irish trees is consistent with this, as most Irish Scots pine are recent reintroductions from Scotland. Under this scenario, the lower within-population relatedness observed in Irish trees can also be expected.

Łabiszak and Wachowiak [108] recently estimated admixture based on variation at 9,760 nuclear SNPs in 62 Eurasian Scots pine populations. Two of these were located in south-central France and both comprised of only one ancestral group when both  $K=3$  and  $K=5$  were inferred. This was the lowest admixture of all studied populations. The higher mean-relatedness in the two French populations sampled here may therefore be explained by Scots pine in this region originating from only one refugial gene pool (Figs. 5 and 6, Figure S10).

Although the Irish populations seem to be part of the same meta-population, some trees located in Derrybawn – Glendalough appeared to be of continental origin (Fig. 6, Figure S10). A mixed origin of this population may account for its relatively high level of genetic variation (Fig. 3). Additional nuclear substructure at  $K=3$  was also observed in two south-western populations (Fig. 6, Figure S10). This was defined by the increased occurrence of a third ancestral group at a  $>50\%$  frequency in all Dale Wood trees, which may reflect the relatively fixed allele frequency in this population, as indicated by



**Fig. 8** Results of a STRUCTURE analysis after removing loci (PtTX3116, SPAC11\_6, psy117, PtTX3107, psy12 and psy125) with excessive null allele frequencies. **a** Mean posterior probability values for  $K$  ( $L(K)$ ), error bars are  $\pm$  S.D. **b** Delta  $K$  values. **c** Bar plots of admixture coefficients at different values of  $K$  for each individual

its high  $F_{ST}$  and low diversity estimates (Figs. 2, 3 and 4, Figure S2). To a lesser extent, several trees in the nearby Torc population also possessed this third group (Fig. 6, Figure S10).

At  $K=2$  and  $K=3$ , Rockforest trees shared genotypic proportions which were indistinguishable from almost all other Irish populations (Fig. 6, Figure S10), most of which were planted between the 19th and 20th century (Table S1). Although this could imply that even Rockforest trees are reintroduced, palynological and macrofossil evidence strongly support an autochthonous origin

of Rockforest pine [30, 33, 39]. Interestingly, at  $K \geq 5$  Rockforest was inferred to harbour genotypic proportions which distinguish it from all other Irish populations (Fig. 8; Figure S12). This differentiation is clearly very minor given the  $F_{ST}$  values for this population are small (Fig. 4). It is therefore unlikely to correspond to any biologically meaningful genetic structure, especially as the overall variation between Irish populations was not great enough for any substructure to be inferred using the spatially explicit clustering approach of TESS3 (Figure S13). However, the subtle shift in allele frequency which

accounts for this signal may nonetheless be the extent to which this relict has become genetically distinct given that its isolation from other native pine (i.e., in Scotland) began only relatively recently. It is also worth noting here that any such signal in nDNA variation may become even more difficult to detect in the future with the effects of homogenising gene flow from reintroduced pine.

#### Is there a genetic link between Irish and Scottish Scots pine?

The hypothesised contribution of Irish Scots pine to Scottish pinewoods is one of the big outstanding questions regarding Scots pine in Ireland and Britain. Some evidence suggests no contribution. For example, the Scottish mitotype which Sinclair et al. [105] suggested to be of Irish origin was subsequently demonstrated to occur throughout Europe in a geography-independent manner [109], which rules out it being an Irish-specific mitotype. Others have argued that the Irish Sea constituted an impermeable barrier to tree migration, with early establishment of pine in England being enough to preclude any significant contribution of Irish pine to Scottish pinewoods [91, 110].

Previous work on *Quercus* species, with much less mobile seed, suggested two colonisation routes into Ireland, from the south west across the Celtic Sea and from the east across the Irish Sea [111]. More recent genetic evidence also suggests at least some post-LGM migration from Britain into Ireland for downy birch (*Betula pubescens* Ehrh.) [47] and common ash (*Fraxinus excelsior* L.) [82]. This corroborated the direction of travel which had previously been inferred from isochrone maps based on radiocarbon-dated pollen for both species [21]. These maps suggest an alternative direction of travel for Scots pine from the south-west of Ireland to the north-east and across to Scotland [21]. Indeed, the Irish Sea has recently been redefined as a “filter” rather than a barrier to seed migration [112, 113], which suggests that migration into Scotland could have occurred. Moreover, wind dispersal of Scots pine pollen is highly efficient, with the potential to travel hundreds of kilometres from southern to northern locations [114, 115]. It is possible that pollen from northern Irish populations could reach Scottish populations, given the prevailing southwest wind direction between Ireland and Scotland, coupled with the smallest distance between both being less than 20 km. Indeed, over millennial timescales, a rare event becomes almost inevitable.

Other works have also alluded to a south-western origin for western Scottish pine. For example, variation at nSSR, isozyme and monoterpene loci in western Scottish pine is more similar to southern rather than northern and central European pine [34, 98, 116]. The absence of

any strong differentiation between Rockforest and reintroduced Scottish Scots pine could potentially be attributable to both originating from the same gene pool. The plantation stock may therefore derive from native Scottish pine which originally migrated from Ireland following the LGM. Resolving this question using molecular markers will be challenging due to the lack of differentiation in Scots pine and uncertainty regarding other suitable sites to test in Ireland. Further insights may be gleaned by sequencing the mitochondrial genome of Rockforest and western Scottish Scots pine, or even sequencing of DNA from preserved pine stumps or fossil pollen elsewhere in Ireland to identify genotypes across a more extended time period. For example, Danusevičius et al. [117] were able to demonstrate an early presence of mitotype *b* in Lithuania based on DNA extracted from 11,000 year-old Scots pine stumps found in the Baltic Sea, while Bennett and Parducci [118] sequenced DNA from 10,000 year-old fossil pollen grains of Scots pine retrieved from lake sediment in Sweden.

#### Conclusions

Whilst Scots pine underwent a regional decline during the Holocene, it is probable that efficient gene flow and a high pre-existing diversity buffered against the species’ terminal decline in Ireland. The results presented here shed some light on the complicated history of Scots pine in Ireland and suggest agreement with the thesis that Irish and Scottish Scots pine share their origins. The lack of clear genetic structure between a putative native Irish population and other populations reintroduced from Scotland suggests that the species across Ireland and Britain exists as a meta-population. This is further underscored by their differentiation from three continental populations and three Norwegian genotypes at nSSR loci. However, more extensive comparison of native Scottish to Irish populations will be needed to confirm this. Whilst it may signal autochthony, the slight shift in allele frequency in Rockforest pine would not seem enough for this population to be treated as genetically distinct from reintroduced pine. Therefore, all Scots pine in Ireland should be managed as if native, as recommended from previous ecological work [8, 39]. Only one population (Derrybawn – Glendalough) harboured trees which appear to be of continental origin, although this was mixed. These results are the first to describe the genetic diversity of a native Irish conifer in a European context and complement recent descriptions of the same for several broadleaf species [47, 82, 119, 120]. This forms part of an inventory of Ireland’s forest genetic resources [121], which is an important step in determining population selection for both *in situ* and *ex situ* forest conservation approaches [122].



## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12862-023-02181-3>.

**Additional file 1.** This file contains scripts used for analysing the data

**Additional file 2: Table S1.** Population names, locations, origins and year of planting, where known, along with haplotype designations and number of individuals per population.

**Additional file 3: Table S2.** Chloroplast microsatellite (cpSSR) primers used to genotype *Pinus sylvestris* individuals in this work. **Table S3.** Nuclear microsatellite (nSSR) primers used to genotype *Pinus sylvestris* individuals in this work. **Table S4a.** Results of SAMOVA analysis, unconstrained by geographic data. **Table S4b.** Results of SAMOVA analysis, constrained by geographic data. **Table S5.** Estimates of haplotype genetic diversity ( $H_{cp}$ ) and expected heterozygosity ( $H_e$ ) based on chloroplast (cpSSR) and nuclear (nSSR) SSR genotyping of *Pinus sylvestris* populations from various regions across its distribution range. Mean values are  $\pm$  S.D. **Table S6.** Estimates of  $H_{cp}$  based on cpSSR loci variation in *Pinus sylvestris*. In each case,  $H_e$  for Irish trees sampled in this work have been re-estimated according to the loci subsets indicated for different works. **Figure S1.** Mean Nei's standard and Bruvo's genetic distance between Scots pine individuals genotyped at cpSSR markers in each sampling population. Error bars are  $\pm$  S.E. **Figure S2.** Pairwise  $F_{ST}$  comparisons between sampled populations. Estimates are based on variation in cpSSR allelic frequency according to Nei (1987). Values which are significantly different from zero are marked with asterisks (\* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ ;  $n = 348$ ). **Figure S3.** Genetic diversity ( $H_e$ ) and allelic richness (rarefaction down to 14 nSSR alleles) for each population. Error bars are 95% CIs derived from 1,000 bootstrap permutations of the data. Estimates in this case were calculated after removing the size loci (PtTX3116, SPAC11\_6, psy117, PtTX3107, psy12 and psy125) displaying null allele frequencies above 10%. **Figure S4.** Pairwise  $F_{ST}$  comparisons between sampled populations. Estimates are based on variation of nSSR allele size differences. Values which are significantly different from zero are marked with asterisks (\* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ ;  $n = 344$ ). **Figure S5.** Mean (a) Nei's standard and (b) Bruvo's genetic distance between Scots pine individuals genotyped at nSSR markers in each sampling population. Error bars are  $\pm$  S.E. **Figure S6.** Distribution of  $r$  estimated under maximum likelihood using nSSR allele variation in sampled Scots pine (*Pinus sylvestris*) trees. **Figure S7.** Distribution of  $r$  estimated under maximum likelihood for each population (including the three and four genotypes from Norway and Scotland (Scottish (Coillte) were samples from a breeding population), respectively) based on nSSR variation in sampled Scots pine (*Pinus sylvestris*) trees. **Figure S8.** Frequency of different pedigree relationships inferred under ML for each population (including the three and four genotypes from Norway and Scotland (Scottish (Coillte)), respectively) based on nSSR variation in sampled Scots pine (*Pinus sylvestris*) trees. **Figure S9.** Principal Coordinates Analysis (PCoA) of the cpSSR allelic composition of each Scots pine (*Pinus sylvestris*) population. Shown are the first and second principal coordinates. Labels are coloured according to population separation in the SAMOVA analysis at  $K=2$ . Coloured shapes are also drawn around these points to assist in interpretation. **Figure S10.** Results of STRUCTURE analysis of Scots pine (*Pinus sylvestris*) nSSR data. In this case, six loci (PtTX3116, SPAC11\_6, psy117, PtTX3107, psy12 and psy125) displaying null allele frequencies above 10% were removed prior to STRUCTURE runs. (a) (i) Mean posterior probability values of  $K$  ( $L(K)$ ), error bars are  $\pm$  S.D. (ii) Delta  $K$  values used to infer the main structure of the data. (b) Frequency of  $K$  groups within each sampling population including those located in Spain, France and Norway (inset). (c) Bar plots of admixture coefficients for each sampled individual. **Figure S11.** Principal Coordinates Analysis (PCoA) of the nSSR allelic composition of each Scots pine (*Pinus sylvestris*) population. Shown are the first and second principal coordinates. Labels are coloured according to the null allele-adjusted STRUCTURE group ( $K=2$ ) which is most frequent in each population. Coloured shapes are also drawn around these points to assist in interpretation. Loci (PtTX3116, SPAC11\_6, psy117, PtTX3107, psy12 and psy125) which displayed null allele frequencies above 10% were removed prior to analysis. **Figure S13.** Results of an analysis of genotype groups ( $K$ ) using TESS3. (a) Given

are the cross-validation (CV) scores based on the root-mean squared errors between the genotypic frequencies predicted from a training set to those computed from a test set (i.e., genotype likelihood) for each locus. (b) The frequency of  $K$  groups within each population are given, including (inset) populations from Spain, France and Norway (for Norway only three individual are used). (c) Individual  $K$  ancestral proportions for each individual. **Figure S14.** Estimates of expected heterozygosity ( $H_e$ ) for *Pinus sylvestris* from three different regions based on 18 nSSR loci (grey) and a reduced set of 12 loci (white). The latter remained after removing six loci (PtTX3116, SPAC11\_6, psy117, PtTX3107, psy12 and psy12) which displayed null allele frequencies above 10%. Error bars are 95% CI values derived from 1,000 bootstrap permutations of the data.

**Additional file 4: Table S7.** Raw cpSSR and nSSR data.

**Additional file 5: Figure S12.** Individual bar plots showing admixture coefficients for  $K = 2$  to  $K = 15$  ancestral groups derived from a STRUCTURE analysis of Scots pine (*Pinus sylvestris*) nSSR variation. Prior to the analysis, six loci (PtTX3116, SPAC11\_6, psy117, PtTX3107, psy12 and psy125) displaying null allele frequencies above 10% were removed.

## Acknowledgements

Funding was received from the CoFoRD Forestry Research Programme of the Department of Agriculture, Food and The Marine for the FORGEN and GeneNet projects. Funding was also received for genotyping from the EU FP7 European Research Infrastructures project Trees4Future. Thanks to Dr. Erwan Guichoux (INRAE) for help with genotyping and Dr. Robledo-Arnuncio (INIA-CIFOR) for kindly providing Spanish DNA samples. We thank all landowners for access to sampling sites, particularly the National Parks and Wildlife Service and Coillte. We are also grateful to Dr. Karsten Hokamp (Trinity College Dublin) for granting access to a high-performance multicore Linux system for running STRUCTURE analyses in parallel.

## Authors' contributions

CK obtained funding and devised the project. PC, SB and CK undertook sampling. PC and SB undertook laboratory and statistical analysis. SB and PC drafted the paper. JR reviewed historical data. All authors edited the final draft of the paper.

## Funding

Funding was received from the CoFoRD Forestry Research Programme of the Department of Agriculture, Food and The Marine for the FORGEN and GeneNet projects. Funding was also received for genotyping from the EU FP7 European Research Infrastructures project Trees4Future.

## Availability of data and materials

The raw cpSSR and nSSR data are available in Table S7 for use in future works.

## Declarations

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare no competing interests.

Received: 22 May 2023 Accepted: 26 November 2023

Published online: 23 January 2024

## References

- Bennett KD. The post-glacial history of *Pinus sylvestris* in the British Isles. Q Sci Rev. 1984;3:133–55.



2. Boratyński A. Range of Natural Distribution. In: Giertych M, Matyas C, editors. Genetics of Scots pine. Development in Plant genetics and Breeding. 1st ed. Amsterdam, Netherlands: Elsevier; 1991. p. 19–30.
3. Carlisle A, Brown AHF. *Pinus sylvestris* L. J Ecol. 1968;56:269–307.
4. Hultén E, Fries M. Atlas of north European vascular plants North of the Tropic of Cancer. Königstein: Koeltz Scientific Books; 1986.
5. McGeever AH, Mitchell FJG. Re-defining the natural range of Scots Pine (*Pinus sylvestris* L.): a newly discovered microrefugium in Western Ireland. J Biogeogr. 2016;43:2199–208.
6. Pyhäjärvi T, Kujala ST, Savolainen O. 275 years of forestry meets genomics in *Pinus sylvestris*. Evol Appl. 2020;13:11–30.
7. Tyrmä JS, Vuosku J, Acosta JJ, Li Z, Sterck L, Cervera MT et al. Genomics of Clinal Local Adaptation in *Pinus sylvestris* under continuous environmental and spatial genetic setting. G3 (Bethesda). 2020;10:2683–96.
8. Roche JR, Mitchell FJG, Waldren S. Plant community ecology of *Pinus sylvestris*, an extirpated species reintroduced to Ireland. Biodivers Conserv. 2009;18:2185–203.
9. Houston Durrant T, de Rigo D, Caudullo G. *Pinus sylvestris* in Europe: distribution, habitat, usage and threats. European Atlas of Forest Tree species. Luxembourg: European Union; 2016. p. e016b.
10. DAFM. Forest Statistics Ireland. Ireland: Department of Agriculture, Food and the Marine; 2022.
11. Bachofen C, Wohlgemuth T, Ghazoul J, Moser B. Cold temperature extremes during spring do not limit the range shift of Mediterranean pines into regions with intermittent frost. Funct Ecol. 2016;30:856–65.
12. Castro J, Zamora R, Hódar JA, Gómez JM. Seedling establishment of a boreal tree species (*Pinus sylvestris*) at its southernmost distribution limit: consequences of being in a marginal Mediterranean habitat. J Ecol. 2004;92:266–77.
13. Vizcaíno-Palomar N, González-Muñoz N, González-Martínez SC, Alía R, Benito Garzón M. Most Southern scots Pine populations are locally adapted to Drought for Tree Height Growth. Forests. 2019;10:555.
14. Naydenov K, Senneville S, Beaulieu J, Tremblay F, Bousquet J. Glacial vicariance in Eurasia: mitochondrial DNA evidence from Scots pine for a complex heritage involving genetically distinct refugia at mid-northern latitudes and in Asia Minor. BMC Evol Biol. 2007;7:233.
15. Pyhäjärvi T, Salmela MJ, Savolainen O. Colonization routes of *Pinus sylvestris* inferred from distribution of mitochondrial DNA variation. Tree Genet Genomes. 2008;4:247–54.
16. Tóth EG, Köbölkuti ZA, Pedryc A, Höhn M. Evolutionary history and phylogeography of Scots pine (*Pinus sylvestris* L.) in Europe based on molecular markers. J for Res. 2017;28:637–51.
17. Parducci L, Jørgensen T, Tollefsrud MM, Elverland E, Alm T, Fontana SL, et al. Glacial Survival of Boreal Trees in Northern Scandinavia. Science. 2012;335:1083–6.
18. Zale R, Huang Y-T, Bigler C, Wood JR, Dalén L, Wang X-R, et al. Growth of plants on the late Weichselian ice-sheet during Greenland interstadial-1? Q Sci Rev. 2018;185:222–9.
19. Huntley B. European post-glacial forests: compositional changes in response to climatic change. J Veg Sci. 1990;1:507–18.
20. Huntley B, Birks HJB. An Atlas of Past and Present Pollen maps for Europe: 0–13000 years ago. Cambridge University Press; 1983.
21. Birks HJB. Holocene isochrone maps and patterns of tree-spreading in the British Isles. J Biogeogr. 1989;16:503–40.
22. Watts WA. Holocene vegetation of the Burren, western Ireland. Lake sediments and environmental history: studies in palaeolimnology and palaeoecology in honour of Winifred Tutin / edited by EY Haworth and JWG Lund. 1984.
23. Mitchell FJG. Where did Ireland's trees come from biology and environment. Proceedings of the Royal Irish Academy. 2006;106B:251–9.
24. Bradshaw RHW, Browne P. Changing patterns in the Post-glacial distribution of *Pinus sylvestris* in Ireland. J Biogeogr. 1987;14:237–48.
25. Mitchell FJG, Maldonado-Ruiz J. Vegetation development in the Glendalough Valley, eastern Ireland over the last 15,000 years. Biology and Environment: Proceedings of the Royal Irish Academy. 2018;118B:55–68.
26. Eronen M, Huttunen P. Radiocarbon-Dated Subfossil pines from Finnish Lapland. Geogr Ann Ser Phys Geogr. 1987;69:297–304.
27. Gear AJ, Huntley B. Rapid Changes in the range limits of Scots Pine 4000 years ago. Science. 1991;251:544–7.
28. Lageard JGA, Chambers FM, Thomas PA. Climatic significance of the marginalization of Scots pine (*Pinus sylvestris* L.) c. 2500 BC at White Moss, south Cheshire, UK. The Holocene. 1999;9:321–31.
29. Little DJ, Mitchell FJG, von Engelbrechten S, Farrell EP. Assessment of the impact of past disturbance and prehistoric *Pinus sylvestris* on vegetation dynamics and soil development in Uragh Wood, SW Ireland. The Holocene. 1996;6:90–9.
30. McGeever AH, Mitchell F. Pine stumps in Irish peats: is their occurrence a valid proxy climate indicator? J Quaternary Sci. 2015;30:489–96.
31. Hannon GE, Bradshaw R, Emborg J. 6000 years of forest dynamics in Suserup Skov, a seminatural Danish woodland. Glob Ecol Biogeogr. 2000;9:101–14.
32. Lust N, Geudens G, Olsthoorn AFM. Scots pine in Belgium and the Netherlands. For Syst. 2000;9:213–31.
33. Roche JR, Mitchell FJG, Waldren S, Stefanini BS. Palaeoecological Evidence for Survival of Scots Pine through the late Holocene in Western Ireland: implications for Ecological Management. Forests. 2018;9:350.
34. Kinloch BB, Westfall RD, Forrest GI. Caledonian scots Pine: origins and genetic structure. New Phytol. 1986;104:703–29.
35. Manning AD, Kesteven J, Stein J, Lunn A, Xu T, Rayner B. Could native scots pines (*Pinus sylvestris*) still persist in northern England and southern Scotland? Plant Ecol Divers. 2010;3:187–201.
36. Sassoon D, Fletcher WJ, Hotchkiss A, Owen F, Feng L. Scots pine (*Pinus sylvestris*) dynamics in the Welsh marches during the mid to late-holocene. The Holocene. 2021;31:1033–46.
37. Roche JR, Mitchell FJG, Waldren S, Bjørndal JE. Are Ireland's reintroduced *Pinus sylvestris* forests floristically analogous to their native counterparts in oceanic north-west Europe? Biology and Environment: Proceedings of the Royal Irish Academy. 2015;115B:97–114.
38. DAFM. Native Woodland Conservation Scheme. Dublin: Department of Agriculture, Food and the Marine; 2015.
39. Roche JR. Recent findings on the native status and vegetation ecology of scots pine in Ireland and their implications for forestry policy and management. Ir Forestry. 2019;76:29–54.
40. Lageard JGA. Pine decline or pine declines? Analysis and interpretation of bog-pines from Wem Moss. Shropsh UK Quaternary. 2023;6:12.
41. Dering M, Baranowska M, Beridze B, Chybicki IJ, Danelia I, Iszkulo G, et al. The evolutionary heritage and ecological uniqueness of scots pine in the Caucasus ecoregion is at risk of climate changes. Sci Rep. 2021;11:22845.
42. Fady B, Aravanopoulos FA, Alizoti P, Mátyás C, von Wühlisch G, Westergren M, et al. Evolution-based approach needed for the conservation and silviculture of peripheral forest tree populations. For Ecol Manag. 2016;375:66–75.
43. Hallingbäck HR, Burton V, Vizcaíno-Palomar N, Trotter F, Liziniewicz M, Marchi M, et al. Managing uncertainty in Scots pine range-wide adaptation under Climate Change. Front Ecol Evol. 2021;9:724051.
44. Julio Camarero J, Gazol A, Sancho-Benages S, Sangüesa-Barreda G. Know your limits? Climate extremes impact the range of Scots pine in unexpected places. Ann Bot. 2015;116:917–27.
45. Reich PB, Oleksyn J. Climate warming will reduce growth and survival of scots pine except in the far north. Ecol Lett. 2008;11:588–97.
46. Roche JR. The Vegetation Ecology and native status of Scots Pine (*Pinus sylvestris* L.) in Ireland. Ph.D. Trinity College Dublin; 2010.
47. Belton S, Cubry P, Fox E, Kelleher CT. Novel post-glacial haplotype evolution in Birch—A case for conserving local adaptation. Forests. 2021;12:1246.
48. Dang XD, Kelleher CT, Howard-Williams E, Meade CV. Rapid identification of chloroplast haplotypes using high Resolution Melting analysis. Mol Ecol Resour. 2012;12:894–908.
49. Soranzo N, Alía R, Provan J, Powell W. Patterns of variation at a mitochondrial sequence-tagged-site locus provides new insights into the postglacial history of European *Pinus sylvestris* populations. Mol Ecol. 2000;9:1205–11.
50. Robledo-Arnuncio JJ, Collada C, Alía R, Gil L. Genetic structure of Montane isolates of *Pinus sylvestris* L. in a Mediterranean Refugial Area. J Biogeogr. 2005;32:595–605.
51. Scalfi M, Piotti A, Rossi M, Piovani P. Genetic variability of Italian southern Scots pine (*Pinus sylvestris* L.) populations: the rear edge of the range. Eur J Forest Res. 2009;128:377.

52. Vendramin GG, Lelli L, Rossi P, Morgante M. A set of primers for the amplification of 20 chloroplast microsatellites in Pinaceae. *Mol Ecol*. 1996;5:595–8.
53. Belletti P, Ferrazzini D, Piotti A, Monteleone I, Ducci F. Genetic variation and divergence in Scots pine (*Pinus sylvestris* L.) within its natural range in Italy. *Eur J Forest Res*. 2012;131:1127–38.
54. Elisk CG, Miniham VT, Hall SE, Scarpa AM, Williams CG. Low-copy microsatellite markers for *Pinus taeda* L. *Genome*. 2000;43:550–5.
55. Elisk CG, Williams CG. Low-copy microsatellite recovery from a conifer genome. *Theor Appl Genet*. 2001;103:1189–95.
56. Sebastiani F, Pinzauti F, Kujala ST, González-Martínez SC, Vendramin GG. Novel polymorphic nuclear microsatellite markers for *Pinus sylvestris* L. *Conserv Genet Resour*. 2012;4:231–4.
57. Zhou Y, Bui T, Auckland LD, Williams CG. Undermethylated DNA as a source of microsatellites from a conifer genome. *Genome*. 2002;45:91–9.
58. Wickham H, Vaughan D, Girlich M. tidy: Tidy Messy Data. 2023.
59. Jombart T. ADEGENET: a R package for the multivariate analysis of genetic markers. *Bioinformatics*. 2008;24:1403–5.
60. Jombart T, Ahmed I. ADEGENET 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*. 2011;27:3070–1.
61. Nei M. Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci U S A*. 1973;70:3321–3.
62. Eliades NGH, Eliades DG. Haplotype Analysis: software for analysis of haplotype data. 2009.
63. Adamack AT, Gruber B. PopGenReport: simplifying basic population genetic analyses in R. *Methods Ecol Evol*. 2014;5:384–7.
64. Gruber B, Adamack AT. Landgenreport: a new r function to simplify landscape genetic analysis using resistance surface layers. *Mol Ecol Resour*. 2015;15:1172–8.
65. Nei M. Genetic Distance between populations. *Am Nat*. 1972;106:283–92.
66. Bruvo R, Michiels NK, D'Souza TG, Schulenburg H. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Mol Ecol*. 2004;13:2101–6.
67. Kamvar ZN, Tabima JF, Grünwald NJ. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*. 2014;2:e281.
68. Kamvar ZN, Brooks JC, Grünwald NJ. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front Genet*. 2015;6:208.
69. Nei M. Molecular evolutionary genetics. New York: Columbia University Press; 1987.
70. Weir BS, Cockerham CC. Estimating F-Statistics for the analysis of Population structure. *Evolution*. 1984;38:1358–70.
71. Goudet J. Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol Ecol Notes*. 2005;5:184–6.
72. Winter DJ. Mmod: an R library for the calculation of population differentiation statistics. *Mol Ecol Resour*. 2012;12:1158–60.
73. Altman DG, Bland JM. How to obtain the P value from a confidence interval. *BMJ*. 2011;343:d2304.
74. Dray S, Dufour A-B. The ade4 Package: Implementing the Duality Diagram for Ecologists. 2007. 2007;22:1–20.
75. Agapow P-M, Burt A. Indices of multilocus linkage disequilibrium. *Mol Ecol Notes*. 2001;1:101–2.
76. Guo SW, Thompson EA. Performing the exact test of hardy-Weinberg proportion for multiple alleles. *Biometrics*. 1992;48:361–72.
77. Paradis E. Pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics*. 2010;26:419–20.
78. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from Incomplete Data via the EM Algorithm. *J Royal Stat Soc Ser B (Methodological)*. 1977;39:1–38.
79. Rousset F. Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour*. 2008;8:103–6.
80. Slatkin M. A measure of Population Subdivision based on microsatellite allele frequencies. *Genetics*. 1995;139:457–62.
81. Chybicki IJ, Burczyk J. Simultaneous estimation of null alleles and inbreeding coefficients. *J Hered*. 2009;100:106–13.
82. Belton S, Fox E, Kelleher CT. Characterising the molecular diversity of ash (*Fraxinus excelsior* L.) at its western marginal range in Europe — phylogeographic insights and implications for conservation in Ireland. *Tree Genet Genomes*. 2022;18:36.
83. Kalinowski ST, Wagner AP, Taper ML. ml-relate: a computer program for maximum likelihood estimation of relatedness and relationship. *Mol Ecol Notes*. 2006;6:576–9.
84. Wagner AP, Creel S, Kalinowski ST. Estimating relatedness and relationships using microsatellite loci with null alleles. *Heredity*. 2006;97:336–45.
85. Dupanloup I, Schneider S, Excoffier L. A simulated annealing approach to define the genetic structure of populations. *Mol Ecol*. 2002;11:2571–81.
86. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945–59.
87. Chhatre VE, Emerson KJ. StrAuto: automation and parallelization of STRUCTURE analysis. *BMC Bioinformatics*. 2017;18:192.
88. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol*. 2005;14:2611–20.
89. Jakobsen M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*. 2007;23:1801–6.
90. Caye K, Deist TM, Martins H, Michel O, François O. TESS3: fast inference of spatial population structure and genome scans for selection. *Mol Ecol Resour*. 2016;16:540–8.
91. Dering M, Kosiński P, Wyka TP, Pers-Kamczyc E, Boratyński A, Boratyńska K, et al. 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. *Divers Distrib*. 2017;23:540–55.
92. Cinget B, Gérardi S, Beaulieu J, Bousquet J. Less pollen-mediated gene flow for more signatures of glacial lineages: congruent evidence from balsam fir cpDNA and mtDNA for multiple refugia in eastern and central North America. *PLoS One*. 2015;10:e0122815.
93. Jiménez-Ramírez A, Grivet D, Robledo-Arnuncio JJ. Measuring recent effective gene flow among large populations in *Pinus sylvestris*: local pollen shedding does not preclude substantial long-distance pollen immigration. *PLoS One*. 2021;16:e0255776.
94. Smith DR. Mutation rates in plastid genomes: they are lower than you might think. *Genome Biol Evol*. 2015;7:1227–34.
95. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A*. 1987;84:9054–8.
96. Wójcikiewicz B, Wachowiak W. Substructuring of Scots pine in Europe based on polymorphism at chloroplast microsatellite loci. *Flora - Morphology, Distribution, Functional Ecology of Plants*. 2016;220:142–9.
97. Provan J, Soranzo N, Wilson NJ, McNicol JW, Forrest GI, Cottrell J, et al. Gene-pool variation in caledonian and European Scots pine (*Pinus sylvestris* L.) revealed by chloroplast simple-sequence repeats. *Proc Biol Sci*. 1998;265:1697–705.
98. Wachowiak W, Salmela MJ, Ennos RA, Iason G, Cavers S. High genetic diversity at the extreme range edge: nucleotide variation at nuclear loci in Scots pine (*Pinus sylvestris* L.) in Scotland. *Heredity (Edinb)*. 2011;106:775–87.
99. Molloy K, O'Connell M. Post-glaciation plant colonisation of Ireland: fresh insights from an Loch Mór, Inis Oirr, Western Ireland. *Ir Naturalists' J*. 2014;33:66–88.
100. Cheddadi R, Vendramin GG, Litt T, François L, Kageyama M, Lorentz S, et al. Imprints of glacial refugia in the modern genetic diversity of *Pinus sylvestris*. *Glob Ecol Biogeogr*. 2006;15:271–82.
101. Semerikov VL, Semerikova SA, Putintseva YA, Tarakanov VV, Tikhonova IV, Vidyakin AI, et al. Colonization history of Scots pine in Eastern Europe and North Asia based on mitochondrial DNA variation. *Tree Genet Genomes*. 2018;14:8.
102. Wachowiak W, Perry A, Zaborowska J, González-Martínez SC, Cavers S. Admixture and selection patterns across the European distribution of Scots pine, *Pinus sylvestris* (Pinaceae). *Bot J Linn Soc*. 2022;200:416–32.
103. Semerikov NV, Petrova IV, Sannikov SN, Semerikova SA, Tashev AN, Lascoux M, et al. Cytoplasmic DNA variation does not support a recent contribution of *Pinus sylvestris* L. from the Caucasus to the main range. *Tree Genet Genomes*. 2020;16:59.

104. Semerikov VL, Putintseva YA, Oreshkova NV, Semerikova SA, Krutovsky KV. Development of new mitochondrial DNA markers in Scots pine (*Pinus sylvestris* L.) for population and phylogeographic studies. *Russ J Genet.* 2015;51:1199–203.
105. Sinclair WT, Morman JD, Ennos RA. Multiple origins for Scots pine (*Pinus sylvestris* L.) in Scotland: evidence from mitochondrial DNA variation. *Heredity.* 1998;80:233–40.
106. Pyhäjärvi T, García-Gil MR, Knürr T, Mikkonen M, Wachowiak W, Savolainen O. Demographic history has Influenced Nucleotide Diversity in European *Pinus sylvestris* populations. *Genetics.* 2007;177:1713–24.
107. Sheller M, Ciocirlan E, Mikhaylov P, Kulakov S, Kulakova N, Ibe A, et al. Chloroplast DNA diversity in populations of *P. sylvestris* L. from Middle Siberia and the Romanian carpathians. *Forests.* 2021;12:1757.
108. Łabiszak B, Wachowiak W. Mid-Pleistocene events influenced the current spatial structure of genetic diversity in Scots pine (*Pinus sylvestris* L.). *Journal of Systematics and Evolution.* 2023. <https://doi.org/10.1111/jse.13013>.
109. Sinclair WT, Morman JD, Ennos RA. The postglacial history of Scots pine (*Pinus sylvestris* L.) in western Europe: evidence from mitochondrial DNA variation. *Mol Ecol.* 1999;8:83–8.
110. Edwards R, Brooks A. The island of Ireland: drowning the myth of an Irish land-bridge? *Ir Naturalists' J.* 2008;29:19–34.
111. Lowe A, Unsworth C, Gerber S, Davies S, Munro R, Kelleher C, et al. Route, speed and mode of oak postglacial colonisation across the British Isles: integrating molecular ecology, palaeoecology and modelling approaches. *Bot J Scotl.* 2005;57:59–81.
112. Harrison S. Never mind the gap: climate, rather than insularity, may limit Ireland's species richness. *Ir Naturalists' J.* 2014;33:107–23.
113. Mitchell FJG. Tree migration into Ireland. *Ir Naturalists' J.* 2008;29:73–5.
114. Koski V. A study of pollen dispersal as a mechanism of gene flow in conifers. *Metsätieteellisen Tutkimuslaitoksen Julkaisuja.* 1970;70:1–78.
115. Varis S, Pakkanen A, Galofré A, Pulkkinen P. The extent of South-North Pollen transfer in Finnish Scots Pine. *Silva Fennica.* 2009;43:717–26.
116. FORREST GI. Relationship of some European Scots Pine populations to native Scottish woodlands based on Monoterpene Analysis. *Forestry.* 1982;55:19–37.
117. Danusevičius D, Buchovska J, Žulkus V, Daugnora L, Girininkas A. DNA markers reveal genetic associations among 11,000-Year-old Scots Pine (*Pinus sylvestris* L.) found in the Baltic Sea with the Present-Day gene pools in Lithuania. *Forests.* 2021;12:317.
118. Bennett KD, Parducci L. DNA from pollen: principles and potential. *The Holocene.* 2006;16:1031–4.
119. Cubry P, Gallagher E, O'Connor E, Kelleher CT. Phylogeography and population genetics of black alder (*Alnus glutinosa* (L.) Gaertn.) In Ireland: putting it in a European context. *Tree Genet Genomes.* 2015;11:99.
120. Perdureau AC, Kelleher CT, Douglas GC, Hodgkinson TR. High levels of gene flow and genetic diversity in Irish populations of *Salix caprea* L. Inferred from chloroplast and nuclear SSR markers. *BMC Plant Biol.* 2014;14:202.
121. Kelleher CT. A National Forest Tree Gene Conservation Strategy and Action Plan for Ireland. *Ir Forestry J.* 2020;77:7–32.
122. de Vries S, Alan M, Bozzano M, Burianek V, Collin E, Cottrell J, et al. Pan-european strategy for genetic conservation of forest trees and establishment of a core network of dynamic conservation units. 2015.

# Publisher's Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

