Journal of Heredity, 2023, **114**, 165–174
https://doi.org/10.1093/jhered/esac062
Advance access publication 4 November 2022
Original Article

Original Article

Population structure and genetic variation of fragmented mountain birch forests in Iceland

Snæbjörn Pálsson[1], Pawel Wasowicz[2], Starri Heiðmarsson[2,3], Kristinn Pétur Magnússon[2,4]

[1]Faculty of Life and Environmental Sciences, University of Iceland, Reykjavík, Iceland,
[2]Icelandic Institute of Natural History, Akureyri, Iceland,
[3]Northwest Iceland Nature Research Centre, Sauðárkrókur, Iceland,
[4]Faculty of Natural Resource Sciences, University of Akureyri, Akureyri, Iceland

Address correspondence to K.P. Magnússon at the address above, or e-mail: kpm@unak.is.

Corresponding Editor: Mark Chapman

Abstract

Betula pubescens Ehrh. (mountain birch) is the only forest-forming tree in Iceland. Since human settlement (874 AD), the continuous 25,000 to 30,000 km^2 forest has shrunk to 1.200 km^2 of fragmented patches, making it a good object to study population genetic consequences of habitat fragmentation and disturbance. Further, genetic studies have also shown that hybridization between the tetraploid ($2n$ = 56) *B. pubescens* and the diploid ($2n$ = 28) *Betula nana* L. (dwarf birch) occurs among Iceland's natural populations. This study assessed the genetic variation within and among 11 birch forests remaining across Iceland. Genotype-by-sequencing methodology provided a total of 24,585 single nucleotide polymorphisms (SNP´s), with a minor allele frequency >5% for genetic analyses. The analysis showed similar diversity within forests, suggesting that fragmentation and hybridization have had a limited effect on the genetic variation within sites. A clear genetic divergence is found among forests from the different regions of Iceland that may reflect historical isolation; the differentiation between forests increased with geographic distances reflecting isolation by distance. Information on the distribution of genetic variation of birch in Iceland is essential for its conservation and to establish genotype–phenotype associations to predict responses to new environmental conditions imposed by climate change and novel biotic/abiotic stressors.

Graphical Abstract

Population structure and genetic variation of fragmented mountain birch forests in Iceland

Pálsson et al., 2022 Journal of Heredity

We defined the genetic structure of remains of *Betula pubescens* forests in Iceland based on genomic variation

Genotype-by-sequencing methodology (GBS) provided a total of 24,585 SNPs, with a minor allele frequency > 5% for genetic analyses

A clear genetic divergence is found among forests from the different regions of Iceland that may reflect historical isolation; the differentiation between forests increased

Key words: conservation, GBS, next-generation sequencing, NGS, SNP: population genetics

Received August 10, 2022; Accepted November 2, 2022

Introduction

Forest fragmentation and habitat loss have caused a significant threat to the diversity of woodlands (e.g. Turner 1996) and genetic diversity within forest species during the last centuries (Schlaepfer et al. 2018). The fragmentation may lead to small isolates of individuals, causing both reduced variation within populations and increased variation among populations, especially if gene flow among the fragmented sites is limited. The lack of variation within populations may have led to depression in presence of inbreeding and limit the population's ability to adapt to environmental changes (e.g. Razgour et al. 2019). On the other hand, the forested patches may have a vital role in restoring the previously forested areas (Anand et al. 2010). Still, their impact will vary depending on the species and how genetically connected the different patches are. For example, a recent study by Borrell et al. (2018) on genome-wide diversity in dwarf birch, *Betula nana*, showed that within the scarce and fragmented populations of Britain, fragments retain sufficient genetic resources for conservation and replanting programs.

Iceland has gone through extensive deforestation since its human settlement around 874 AD. The forest coverage has been reduced from 20% to about 1.5% of the total area of Iceland (103,000 km^2) and is currently represented by fragmented patches of birch forests scattered throughout the country (Aradottir and Eysteinsson 2005) (Fig. 1). The downy birch (*Betula pubescens* Ehrh.) is a tetraploid species and the only native tree that has dominated Icelandic forests during the Holocene (Hallsdóttir 1995). Conservation of the remaining birch forests is a recognized priority in Iceland. The main tree-like birch taxon in Iceland has been classified as *B. pubecsens* subsp. *tortuosa* (*Ledeb*) Nyman (Wasowicz 2020), together with shrub-like birches in sub-Alpine regions of Fennoscandia and the eastern Kola Peninsula (Walters 1964; Jónsson 2004). Although the forests in Iceland may have been continuously distributed over large areas, certain areas have been isolated for a long time due to the physical barrier of glaciers, sands, and mountains, as is the case in southeast Iceland. Most of the country's interior is above 400 m and covered with sparse or no vegetation, with few notable exceptions. In recent decades, birch growth and distribution have increased due to more favorable conditions such as increased temperatures and diminished sheep grazing, which decreased sharply from 1980 to 2000 (Arnalds and Barkarson 2003). Icelandic birch forests have also benefitted from reforestation and further protection. Whether this fragmentation has affected the genetic variation within the species is unclear. *B. pubescens* and other birch species have a large capacity for gene flow through airborne pollen. In Iceland, pollen from mainland Europe is recorded in early spring, which may be dispersed among local trees (Przedpelska-Wasowicz et al. 2021). Long-distance gene flow from core areas may introduce crucial raw material for adaptation but may also have maladaptive effects in peripheral populations such as those in Iceland (Kremer et al. 2012).

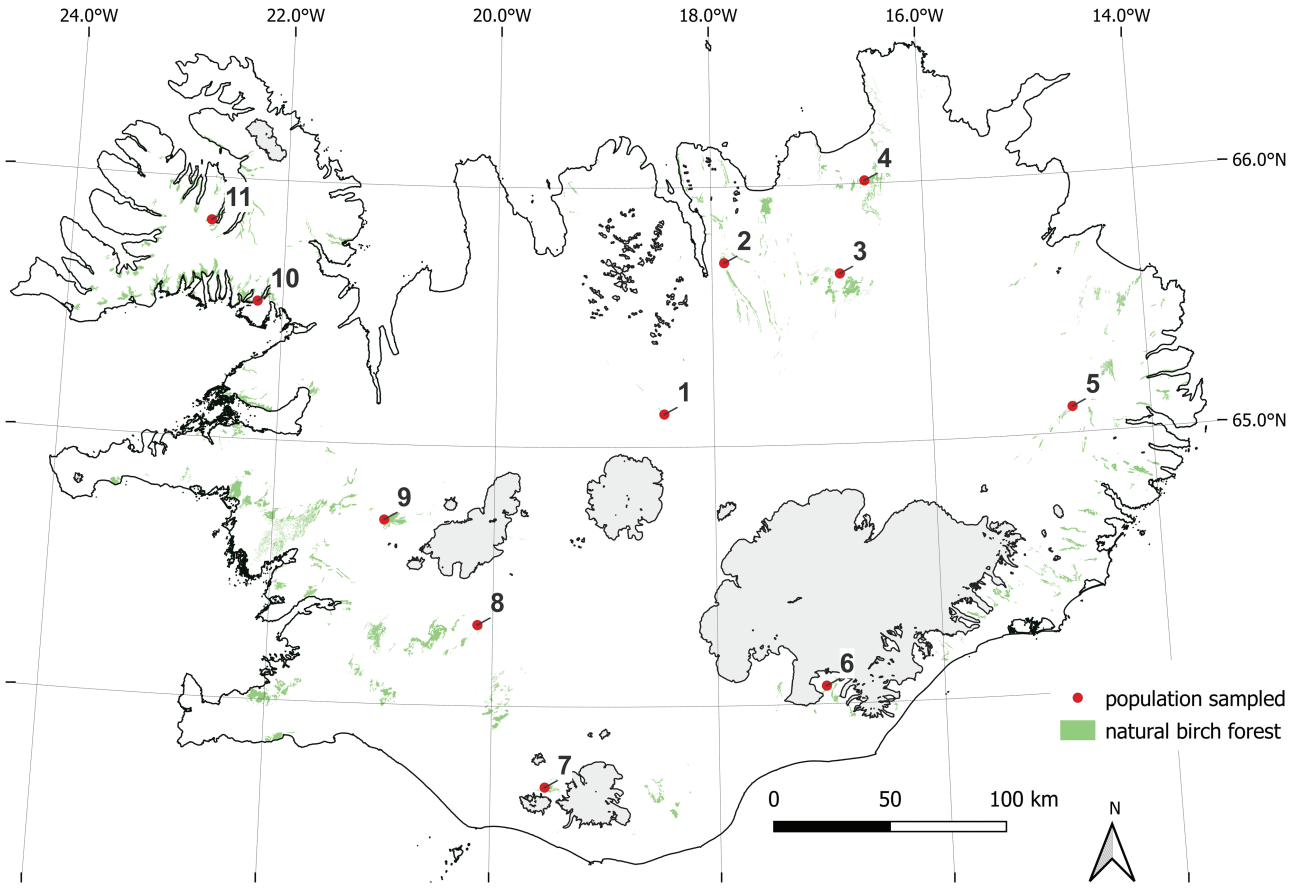

Fig. 1. Sampling sites of *Betula pubescens* in Iceland. Further information on sampling sites 1 to 11 are given in Table 1.

Table 1. Samples across Iceland (I).

Label	Area	Locality	Lat; Long	Date
1	N	Fagrahlíð	65.127; –18.457	18.08.2020
2	N	Vaglaskógur	65.707; –17.885	03.07.2020
3	N	Hlíðardalur	65.656; –16.805	03.07.2020
4	N	Meiðavallaskógur	66.011; –16.542	03.07.2020
5	E	Hallormsstaðaskógur	65.100; –14.718	20.07.2020
6	SE	Bæjarstaðaskógur	64.070; –17.045	26.03.2019
7	S	Þórsmörk	63.689; –19.526	9.07.2020
8	S	Tungufellskógur	64.312; –20.136	8.07.2020
9	W	Skógarhraun	64.710; –20.993	10.07.2020
10	WF	Teigsskógur	65.533; –22.235	20.07.2020
11	WF	Heydalur	65.836; –22.703	22.07.2020

Nine individuals were sampled at each site, except at Bæjarstaðaskógur where 6 individuals were sampled. Areas are north (N), east (E), southeast (SE), south (S), west (W), and Westfjords (WF).

Studies by Thorsson et al. (2007) and Thórsson et al. (2010) on variation in *B. nana* and *B. pubescens*, the 2 native *Betula* species in Iceland, showed extensive morphological variation and frequency differences of chloroplast variants among populations sampled throughout the country. The studies reported a widespread hybridization between the 2 species and a greater extent of introgression between the 2 species in Iceland than in samples from northern Europe (Palmé et al. 2004) and which was also higher in marginal, open or disturbed grounds than in woodlands with dense vegetation (Thorsson et al. 2007). The study by Thórsson et al. (2010) showed furthermore that the introgression was negatively associated with the geographic distances between samples of the 2 species, suggesting that the hybridization was recent. Although the analysis based on chloroplast variation indicated bidirectional gene flow, recent work based on RadSeq analysis of nuclear genomic variants in Britain has revealed a directional gene flow from the diploid *B. nana* and *B. pendula* Roth toward the tetraploid *B. pubescens* (Zohren et al. 2016). High genetic variation can also be expected in *B. pubescens*, especially if it is an allopolyploid species (Salojärvi et al. 2017); that may have arisen from hybridization of 2 distinct species, possibly by *B. pendula* and *B. humilis* Schrank (Tsuda et al. 2017), but polyploidy maintain also more diversity within a population due to limited genetic drift.

The main aim of this study was to explore the genetic structure of *B. pubescens* in Iceland based on genomic variation. More specifically, to evaluate whether fragmentation of the indigenous *B. pubescens* forest has led to genetic differentiation among forest patches and reduced genetic variation within them and whether the putative differentiation will follow geographic distances. Based on the variation within the species we will also consider whether we detect any distinct genotypes or abrupt population differences as 1 might expect by ongoing local hybridization between *B. pubescens* and *B. nana* or long-distance dispersal of pollen or seeds from distantly related individuals of *B. pubescens* from mainland Europe.

Methods

Samples

Leaves were collected from 96 individual *B. pubescens* trees from 11 forests across Iceland in the years 2019–2020 (Fig. 1 and Table 1). Four forests were sampled in north Iceland (1:4), 1 in the east (5), 1 in Bæjarstaðaskógur in the southeast (6), 2 in the south (7 and 8), 1 in the west (9), and 2 from Westfjords (10:11) (Table 1). Each sample consisted of 9 individuals, except sample nr 6 where 6 individuals were analyzed.

The trees selected for this study were larger than 2 m, all GPS positioned and photographed. The leaves were dried in the field with silica gel and stored at room temperature. Samples were sent to LGC (Berlin, Germany), where DNA was extracted and libraries prepared for genotype-by-sequencing (GBS), using the restriction enzymes *Pst*I and *Ape*KI, and sequenced on Illumina NovaSeq SP FC. Bioinformatic processing of the sequence reads was done at LGC using freebayes-v1.2.0 (Garrison and Marth 2012) and Bowtie (Langmead et al. 2009) as:

```
freebayes --min-base-quality 10 --min-supporting-allele-qsum 10 --read-mismatch-limit 3 --mincoverage 5 --no-indels --min-alternate-count 4 --report-genotype-likelihood-max --exclude-unobserved-genotypes --genotype-qualities --ploidy 4 --min-alternate-fraction 0.0833333333333 --no-mnps --no-complex --mismatch-base-quality-threshold 10 --fasta-reference all_joinedSR_clipped_passed-re-filter_singleton-filtered-unique_clusters_min-total-count-20_min-len64.fasta --bam All_0_sorted.bam –targets All_batch0.bed.
```

The genotyping generated a dataset of 177,688 SNPs, with 24,585 SNPs at minor lower allele frequency (MAF) >0.05, in 10,894 contigs. Filtering of variants was carried out by a GBS-specific ruleset where the read count for a locus exceeded 8 reads, and genotypes were observed in at least 66% of each sample. Diploid genotypes were in low frequency (1.9% out of total), possibly due to allele droppage or null alleles, and were excluded from the statistical analysis. This lowered the number of contigs to 6,559.

Statistical analysis

The genomic variation was analyzed based on different sets of data (hereafter referred to as datasets 1, 2 and 3): 1) all markers genotyped (24,585 SNPs), 2) for the MAF of 0.05, sampling 1 SNP per contig to avoid linked variants (6,559

Table 2. Variation within samples.

Label	Variance	H_E	*Ho*	F_{IS}	Singletons
1	0.0141	0.389	0.351	0.099	0.034
2	0.0136	0.383	0.349	0.088	0.031
3	0.0138	0.387	0.354	0.086	0.029
4	0.0142	0.387	0.356	0.081	0.028
5	0.0139	0.377	0.343	0.089	0.030
6	0.0150	0.368	0.344	0.064	0.032
7	0.0139	0.383	0.350	0.084	0.036
8	0.0147	0.396	0.357	0.097	0.045
9	0.0148	0.386	0.352	0.088	0.026
10	0.0139	0.395	0.361	0.087	0.033
11	0.0136	0.399	0.364	0.087	0.037

Variance was estimated based on Kosman distances. Expected (H_E) and observed (H_O) present the proportions of heterozygous genotypes, with doses 1, 2, or 3, averaged over loci. F_{IS} presents the inbreeding coefficient. Heterozygosity and F_{IS} are based on markers with MAF >0.05, complete genotypes and a single SNP per contig. Singletons present a proportion of alleles that are unique to the sample.

SNPs), and 3) MAF ≥05, 1 SNP per contig and only for markers with complete genotypes for all individuals (3,791 SNPs), as required by Discriminant Analysis of Principal Components (DAPC) (Jombart et al. 2010, see below).

The genetic information variation of all markers (dataset 1) was summarized by Kosman distances (Kosman and Leonard 2005), known to be useful for tetraploid individuals, based on the proportion of the nucleotides at each variable site. Variation at each location was summarized by calculating the variance among individuals as $\mathrm{sum}(d_{ij}^2)/(n(n-1))/2$, where d_{ij} is the Kosman distance between individuals i and j and n is the sample size. Variation in Kosman genetic distances (dataset 1) among samples (Φ_{ST}) was summarized with AMOVA, using the R-package pegas (Paradis 2010), and pairwise comparisons were summarized with net Kosman genetic distances ($D_n = D_{XY} - (D_{XX} + D_{YY})/2$), where D_{XY} is the average distance between locations X and Y, and D_{XX} and D_{YY} within locations X and Y. The statistical significance of the analysis of Kosman distances was tested with 1,000 replicates of permutation of individuals among populations. Ordination of individuals based on the Kosman genetic distances was visualized using a multidimensional scale plot (using the cmdscale function in R) and by calculating the average scores for individuals from each site. Isolation by distance was evaluated by comparing the net genetic distances (D_n) and geographic distances between the locations and tested with a Mantel test in the R-package vegan (Oksanen et al. 2014). The association of the net genetic distances and latitudes and longitudes was also tested with the adonis function in vegan (an ANOVA-permutation like test). The number of unique alleles per sample (singletons) and which were shared by more than 1, were summarized and considered with respect to the order of the samples around the country.

Variation at individual SNPs (datasets 2 and 3) was summarized with the R-package vcfR (Knaus and Grünwald 2017) or otherwise as listed using own script in R (R Core Team 2020). Observed heterozygosity was calculated for each individual and location as 1-proportion of sites per individual with homozygous SNP variants (dose 0 or 4 i.e. the frequency of the reference allele per SNP 0: 0/0/0/0, 1: 0/0/0/1, 2: 0/0/1/1, etc.). Expected heterozygosity at each locus within location was calculated as $1 - \mathrm{sum}\left(p_i^4\right)$, and for each dose (0 to 4) assuming binomial probabilities. The expected and observed heterozygosity (H_E and H_O) deviation for all loci per location was summarized with average F_{IS}, and evaluated separately for doses 1, 2, and 3. The partition of variance of the SNP variants among populations and between all was calculated as $F_{ST} = 1 - H_S/H_T$ (also known as G_{ST}), where H_S and H_T are the average and total heterozygosities. A confidence interval for the overall F_{ST} was obtained with bootstrapping over loci, 10,000 times and from the 2.5% and 97.5% quantiles.

The robustness of the analysis was evaluated by comparing the results obtained from the 3 datasets. The correlation of the F_{ST} distances matrices was conducted with the Mantel test (Oksanen et al. 2014).

Clustering of genotypes, based on the dose of 0 to 4 nucleotide per loci, and assignment of individuals to samples based both on prior information of sampling sites and solely based on genetic similarities were done using the DAPC method (Jombart et al. 2010), implemented in the adegenet package in R. This was done only for loci with complete genotypes (dataset 3), i.e. with no missing loci. The number of genetic clusters was selected based on the lowest BIC score.

To assess the putative signal of hybridization observed in chloroplast DNA, the chloroplast DNA sequence from *B. pubescens* (genbank accession nr NC_03996.1, Yang et al. 2019), of unknown geographic origin, was blasted (using BLASTN 2.9.0+, Zhang et al. 2000) against the GBS reads to assess its occurrence within the data. The fragments identified were then blasted (using nblast at https://www.ncbi.nlm.nih.gov/) against the available *B. nana* chloroplasts DNA sequences in genbank, one of unknown origin (KX703002.1, Hu et al. 2017), and 7 from Russia (MT872524.1: MT872570.1, Meucci et al. 2021).

Results

Variation was similar within samples. Variance in Kosman genetic distances per sample ranging from 0.0136 to 0.0150, and averages of expected heterozygosities H_E ranging from 0.367 to 0.394 in dataset 3 (Table 2), the range of H_E was similar

in dataset 2, ranging from 0.376 to 0.401. No individuals deviated clearly from the others within each location, neither when considering the Kosman distances, ranging from 0.16 to 0.18 within locations, or in observed heterozygosity H_O per individual which had a unimodal distribution (Fig. 2) with a mean of 0.348 and SD = 0.011 (dataset 3), and somewhat less for dataset 2 (mean = 0.304 SD = 0.010). Slightly fewer heterozygotes were observed than expected, as reflected by the inbreeding coefficient (F_{IS}), which ranged from 0.0645 to 0.103 for dataset 3 (Table 2), higher deviation was seen for dataset 2 due to lower H_O (0.193 to 0.228). The deviation in F_{IS} for dataset 3 is due to fewer genotypes of dose 1 or 3, than expected, ranging from 0.119 to 0.170), for dose 2 there was slight excess, F_{IS} ranged from –0.117 to –0.050. The proportion of unique alleles or singletons found in each population ranged from 0.026 to 0.045.

The overall genetic differentiation among samples was small but significant but was less for Kosman distance (Φ_{ST} = 0.024, $P < 0.001$) than for F_{ST} = 0.044, 95% CI = 0.043 to 0.045 (dataset 3) and F_{ST} = 0.057, 95% CI = 0.056 to 0.058 (dataset 2). Despite small differentiation, clear differentiation was seen in pairwise comparisons for both D_n and F_{ST}, distances (Table 3 and Supplementary Table 2). The pairwise F_{ST} values for datasets 2 and 3 were highly correlated (r = 0.988, $P < 0.001$, Mantel test, 1,000 permutations) and thus we only describe the result for F_{ST} for dataset 3. Weaker, and a nonsignificant correlation was observed between D_n and F_{ST} (r = 0.35, P = 0.094), but both distances were positively correlated with geographic distances, D_n (r = 0.61, $P <$ 0.001, Mantel test, 1,000 permutations, Fig. 3) and F_{ST} (r = 0.41, P = 0.02). The D_n distances increased both with respect to latitude and longitude (Table 4). Although the D_n and F_{ST} were not significantly correlated they showed overall similar patterns and location of samples in a multidimensional scale plot follows the circular pattern of the sampling locations within Iceland both for D_n (Fig. 4) and F_{ST} (Supplementary Fig. 1). Samples 5 and 6 from east are most divergent from the other although the small sample from Bæjarstaðaskógur (6) did often not deviate significantly from the others: The samples within regions cluster together, i.e. the samples in the north (2, 3, and 4), the ones in south and west (7, 8, and 9) and the 2 from Westfjords (10 and 11). Ordination of sample 1 in northwest varied between the plots for the D_n and F_{ST}, it was closest in the former to the northern samples but whereas for the F_{ST} it was closest to the 3 in south and west.

Clustering of genotypes and assignment probabilities either to each location or to genetic clusters showed a similar result as obtained with the analysis of the distances (Fig. 5); geographically adjacent samples were more alike than those sampled further away. When geographic information was taken into account in the calculations, samples

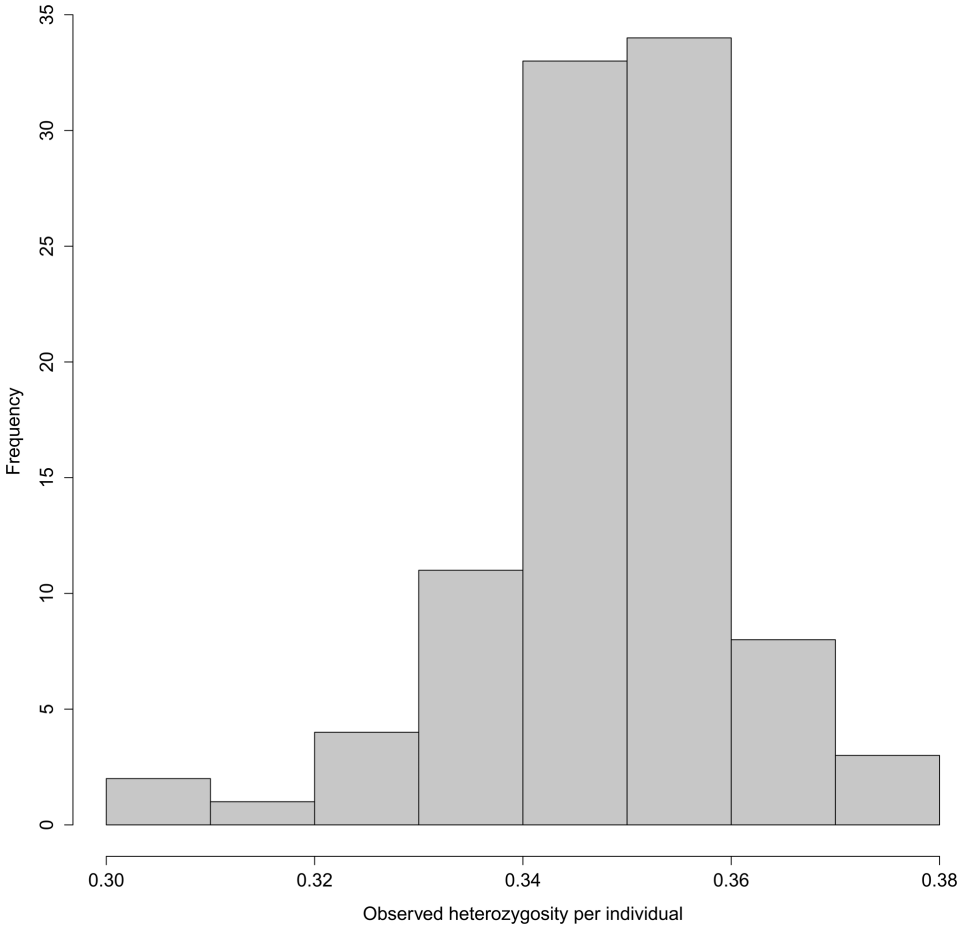

Fig. 2. Distribution of the observed heterozygosity per individual of *Betula pubescens* from Iceland. The estimates are obtained from 1 SNP per 3,791 contigs and complete genotypes.

Table 3. Pairwise net genetic distances (D_{xy}) between samples across Iceland.

	1	2	3	4	5	6	7	8	9	10	11
1	—	0.002	0.001	0	0.005	0.002	0.002	0.001	0.001	0.002	0.002
2	0.01	—	0.001	0.001	0.006	0.001	0.002	0.003	0.002	0.003	0.004
3	0.40	0.59	—	0	0.004	0.001	0.003	0.002	0.001	0.002	0.004
4	0.84	0.18	1.0	—	0.002	0.001	0	0.001	0.001	0.001	0.002
5	0	0	0	0	—	0	0.004	0.005	0.004	0.006	0.006
6	0.06	0.36	0.12	0.15	0.87	—	0.001	0.001	0.002	0.004	0.004
7	0.02	0.01	0	0.59	0	0.07	—	0	0	0.002	0.004
8	0.02	0	0	0.08	0	0.24	1	—	0.001	0.003	0.003
9	0.31	0	0.01	0.04	0	0.01	1	0.12	—	0.001	0.001
10	0.02	0	0	0.02	0	0	0	0	0.15	—	0.001
11	0	0	0	0	0	0	0	0	0.02	0.46	—

The distances are above the diagonal, *P* values below the diagonal were obtained with 100 permutations.

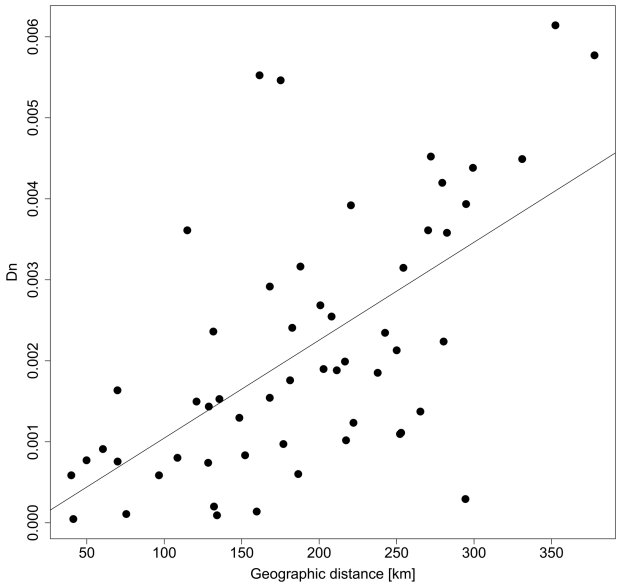

Fig. 3. Association of genetic and geographic distances. Genetic distances are summarized with net genetic distances (D_a) of the Kosman distances between sample sites. A line from a linear regression is drawn ($r = 0.61$, $P < 0.001$, Mantel test, 1,000 permutations).

1 and 2 in north Iceland clustered together and in close vicinity to the other 2 samples from the north (3 and 4) but also to site 9 in west Iceland (Fig. 5a). Sites in the east (5) and southeast (6) were well separated from the other to the left, and also the similar genotypes in 10 and 11 from the Westfjords. When no prior information was used (Fig. 5b) similar patterns were observed but the samples grouped into 4 clusters. Clusters 1 and 2 were composed of individuals from south, west, and north of Iceland, where cluster 1 was mainly from the north and cluster 2 from the others. Cluster 3 was characterized by individuals from the Westfjords and cluster 4 from the distinct samples from east (sites 5 and 6). A more detailed assignment probabilities of the groups are shown in Fig. 5c and d. The observed number of clusters in Fig. 5b and d was based on the lowest BIC value obtained when no geographic information was considered (Supplementary Fig. 2).

Almost all individuals (90%) were assigned to their sampling site with assignment values ranging from 0.78 to 1.0. When an individual had a higher assignment to another site than the site of origin, it was to the neighboring sampling site (Fig. 6). One individual at site 1 had higher assignments to site 9, than to its sampling site whereas 2 individuals at site 9 had higher assignment to sample 1; 1 individual at site 3 had highest assignments to site 4. Lastly, for both sites 10 and 11, 2 individuals had the highest assignment to the other respective site (Fig. 5c).

The proportion of alleles, shared by only 2 populations, summarized for dataset 2, decreased with the sampling order around the country ($r = -0.31$, $P < 0.01$). The nearest neighboring samples shared on average 25% of alleles not found elsewhere, whereas pairs separated by 1 to 4 samples around the island shared 17% to 19% of alleles.

Eight contigs showed evidence of chloroplast origin (*E*-value ranging from 1×10^{-133} to 4×10^{-08}), but 5 of them matched only partly the chloroplast sequence; the ratio of the chloroplast read and the total contig length was 92/239, 70/243, 48/156, 75/191, and 34/146, suggesting that these reads are numts. The whole sequence of 3 contigs showed a match to the birch chloroplast (259/259, 175/175, and 174/176). In total these reads were of a length of 926 bases but unfortunately no difference was found between *B. nana* (KX703002.1) and *B. pubescens* (NC_03996.1) for these sequences, but some variance was observed in the *B. nana* sequences from Russia and 2 of the variable sites were shared with variable sites in Iceland.

Discussion

Genomic variation within *B. pubescens* in Iceland is similar within locations but differs between areas. The differentiation between sites follows an isolation by distance model where neighboring populations are more alike than those separated by larger geographic distances, reflecting limited dispersal. The differentiation between populations is though small and may be recent as most of the variation is within populations. The fragmentation of the birch forest since human settlement seems to have had smaller impact on this pattern than 1 might expect e.g. where certain small forest remains could deviate significantly from their neighboring sites due to increased

Table 4. Analysis of variation in genetic distances with respect to latitudes and longitudes.

Source	df	Mean Sq	F'	R^2	$P(>F)$
Latitude	1	0.0206	1.445	0.015	0.001
Longitudes	1	0.0272	1.911	0.020	0.001
Residuals	93	0.0142		0.965	

The significance was estimated with a permutational ANOVA (999 permutations) in adonis (Oksanen et al. 2014).

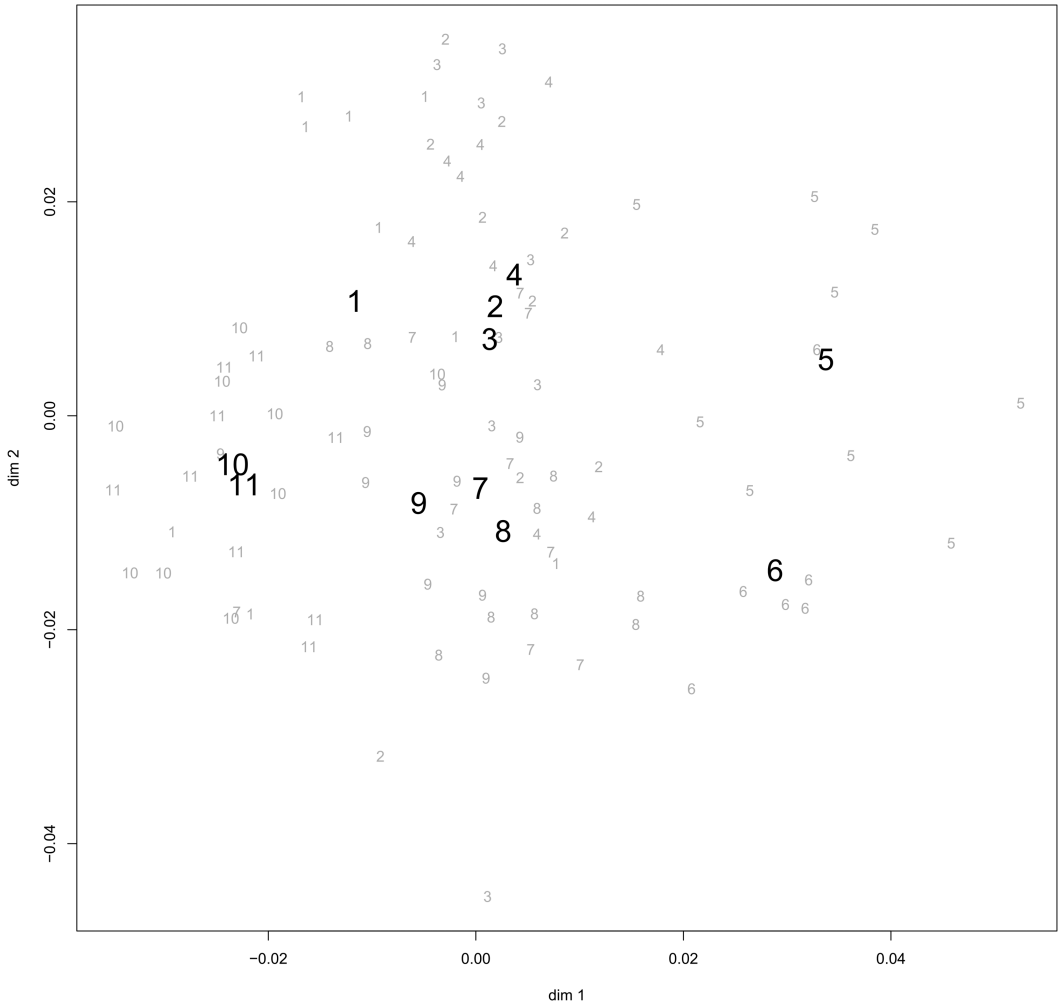

Fig. 4. Multidimensional scale plot of distances between birch genotypes. The plot is based on Kosman distances between individuals numbered by their geographic origin 1:11 (see Fig. 1 and Table 1). The average of coordinates for each sample is given by bigger numbers in black.

genetic drift. Similar result was also found in a study on *B. nana* in Scotland (Borrell et al. 2018), but the overall variation within Iceland may be considered to be lower today than before the deforestation of the country. Assignment of individuals to locations supported this differentiation as well as the distribution of rare alleles, but neighboring populations were more likely to share rare alleles than sites further away. Sites with fewer geographical barriers or where forests may have been continuous over larger areas such as in south, west, and north Iceland are less differentiated than the more geographically isolated sites; the east and southeast Iceland sites are isolated by the Vatnajökull glacier, highlands, and glacial rivers, the sites in the Westfjord-peninsula are isolated by large distances, mountains, and the ocean.

A study on chloroplast variation by Thórsson et al. (2010) indicated that hybridization between *B. nana* and *B. pubescens* was widespread in Iceland and that plants from the same area were more likely to share the same chloroplast than plants further away. Here, we observe similar diversity within all samples and isolation by distance but no distinct outliers or abrupt changes in diversity or divergence of populations as might be expected if hybridization or long-distance migration has taken place e.g. from mainland Europe. Despite extensive introgression of chloroplast among North American birches, they have maintained their nuclear separation as detected with microsatellites (Thomson et al. 2015). A clear delimitation between *Betula* species in Eurasia has also been observed with nuclear simple sequence repeat (SSR) markers, although

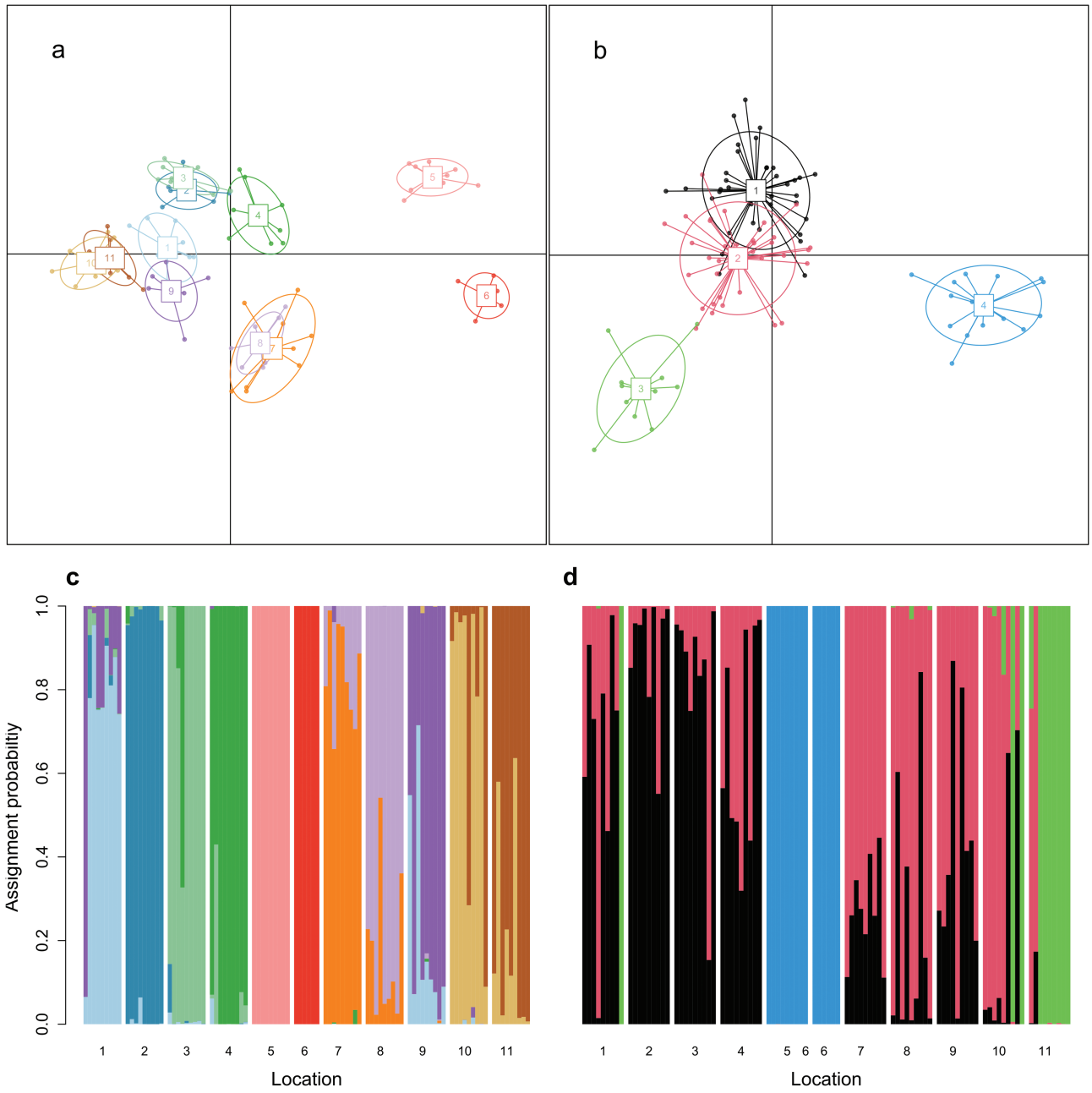

Fig. 5. Clustering of individuals based on genetic similarities and membership probabilities. Prior information of sampling origin was used in (a), no prior information was used in (b) which was based on the number of clusters ($K = 6$) supported by the lowest BIC (see Supplementary Fig. 1). Assignment plots, presenting membership probabilities to each site, based on the 2 scenarios with and without prior geographic information, are presented in (c) and (d).

several cases of introgression were observed (Tsuda et al. 2017). Thus, the high introgression rate in Iceland, observed by Thórsson et al. (2010), appears to be limited to the chloroplast, although analyses including GBS of *B. nana* is needed to assess putative introgression in the nuclear chromosomes of the 2 species but difference in sampling of individuals may also explain the different outcome in this study. To understand the magnitude of hybridization fully, the coverage and coexistence of the 2 species, the characteristics of the hybrids (morphology and/or ploidy level) as well as the habitat which may drive the hybridization (Thorsson et al. 2007), needs to be considered and the plants sampled accordingly. *B. nana* is more evenly distributed countrywide with the exception of South Iceland, including Þórsmörk and Bæjarstaðaskógur (samples 6 and 7), where it is missing while *B. pubescens* has a more patchy distribution (Ottósson et al. 2016, Snorrason et al. 2016). The low level of diversity among the samples in our study may reflect that *B. pubescens* is self-incompatible, with both its seeds and pollen dispersed by wind (Kremer 1994; Hamrick and Godt 1996; Austerlitz et al. 2000). However, despite low differentiation between forests, the differentiation follows the geographic distances, and neighboring sites share unique alleles, thus suggesting that dispersal is limited or more common among neighboring sites than over long distances. The lack of deviant genotypes also indicates that dispersal by seeds or gene flow via pollen from other countries may

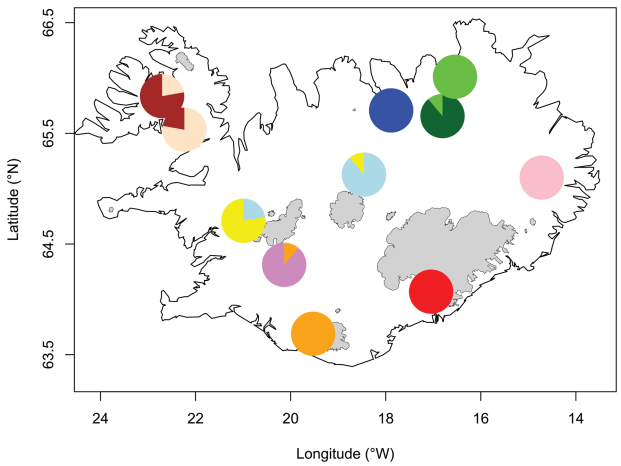

Fig. 6. Summary of the assignments of individuals to geographic sites. Colors in pies present individuals' most likely geographic origin based on the assignment probability from the DAPC analyses with prior information of the geographic sampling site, as presented in Fig. 5.

be uncommon. Pollen in Iceland originating from mainland Europe peaks earlier during the year than the flowering season of Icelandic birches (Przedpelska-Wasowicz et al. 2021) and may thus have a limited impact on the gene pool in Iceland. Analyses of chloroplast haplotypes by Thórsson et al. (2010) and comparison with corresponding haplotype frequencies in mainland Europe (Palmé et al. 2004) showed that the haplotype frequencies were most like samples from northern Scandinavia. As seeds mainly disperse the chloroplasts, this may reflect shared ancestry rather than recent or ongoing gene flow. Tsuda et al. (2017) also showed similar patterns in SSR in samples from Iceland and northern Europe. Whether slight genetic differentiation within Iceland emerged due to different colonization routes after the last glaciation remains to be answered with a further genomic comparison with the birch from the British Isles and Scandinavia.

Attempts of reforestation in Iceland should take the results of this study into account. In order to preserve the genetic diversity of *B. pubescens*, seeds used for replantation could be sampled within their respective regions. The seeds may harbor locally adapted gene pools which may have diverged from the most distinct samples since the settlement of birch in Iceland about 8,000 years BP. Trees from Bæjarstaðaskógur in Southeast Iceland have been favored due to their morphological characteristics (pillar with a single stem), and they have also been shown to have less chloroplast introgression from *B. nana* (Thórsson et al. 2010). Whether they should be used as a seed source for other areas is an open question and it is clear that source populations should be selected with care to maintain genetic diversity. The genetic distances between sites reported here are subtle, and thus the morphological differentiation may be associated with few functional genes which in turn might depend on environmental factors. In a recent review, Hill and Hollender (2019) presented discoveries on the regulation of shoot architectures for which causative genes have been identified, including dwarf, weeping, columnar, and pillar growth habits. The hybridization and gene flow between the 2 species in Iceland are ideal for a study on comparative genomics and to assess the impact of introgression from *B. nana* on the morphology of *B. pubescens*. Furthermore, due to polyploidy this species might harbor genome differentiation and rearrangements which might contribute to the genetic diversity (Wang et al. 2021) and the variable growth forms within the species in Iceland. In addition to the impacts on morphology, characterization of the *B. pubescens* genome is also crucial to reveal the genetic components derived from its putative ancestors, *B. pendula* and *B. humilis* (Tsuda et al. 2017). This characterization could similarly be used to evaluate whether introgression from *B. nana* has affected the growth of the *B. pubescens* differently in the fragmented forests in Iceland, and possibly less so in Bæjarstaðaskógur.

Climate change is the main driver of major environmental changes across the whole Arctic region (Taylor et al. 2020) that is manifested, inter alia, by increased activity of invasive alien species (Wasowicz et al. 2020). In addition to concerns related to the morphological structures discussed above, emerging threats for Icelandic birch are new (most probably imported) leaf-mining insects: *Scolioneura betuleti* and *Heringocrania unimaculella* that can cause significant leaf damage and defoliation (Tomczyk et al. 2017). Even though the abundance and distribution of genetic resistance to these insects are currently unknown, the maintenance of broad genetic diversity has been shown to be essential in coping with newly introduced pests and pathogens (Budde et al. 2016).

In summary, the patterns are clear: 1) similar variation is observed at all sites, 2) isolation by distance indicates limited dispersal, despite the dispersal capacity. That is to say, neighboring populations are more similar to each other than those further apart—i.e. the pollen is not able to override the local differentiation, mainly because of the difference in the phenology of *B. pubescens* in Iceland and Europe. This is important both for management/conservation and for adaptation to local conditions (Oksanen 2021).

Supplementary material

Supplementary material is available at *Journal of Heredity* online.

Funding

This work was supported by the University of Akureyri Research Fund (nr. R2007) and The Nature Conservation Fund of Pálmi Jónsson.

Acknowledgments

We thank Guðný Vala Þorsteinsdóttir at IINH for preparing the leave samples for shipment. We also thank Roza Parol-Kryger at LGC for valuable representation of the data.

Conflict of interest

The authors declare no conflict of interest.

Data availability

The genotype dataset associated with this paper is accessible at https://doi.org/10.5281/zenodo.7230256.

References

Anand MO, Krishnaswamy J, Kumar A, Archana B. Sustaining biodiversity conservation in human-modified landscapes in the Western

Ghats: remnant forests matter. *Biol Conserv*. 2010;143:2363–2374.
Aradottir AL, Eysteinsson T. Restoration of birch woodlands in Iceland. In: Stanturf JA, Madsen P, editors. *Restoration of boreal and temperate forests*. Boca Raton (FL): CRC Press; 2005. p. 195–209.
Arnalds O, Barkarson B. Soil erosion and land use policy in Iceland in relation to sheep grazing and government subsidies. *Environ Sci Policy*. 2003;6:105–113.
Austerlitz F, Mariette S, Machon N, Gouyon PH, Godelle B. Effects of colonization processes on genetic diversity: differences between annual plants and tree species. *Genetics*. 2000;154:1309–1321.
Borrell JS, Wang N, Nichols RA, Buggs RJA. Genetic diversity maintained among fragmented populations of a tree undergoing range contraction. *Heredity*. 2018;121:304–318.
Budde KB, Nielsen LR, Ravn HP, Kjær ED. The natural evolutionary potential of tree populations to cope with newly introduced pests and pathogens—lessons learned from forest health catastrophes in recent decades. *Curr For Rep*. 2016;2:18–29.
Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv*. 2012. doi:10.48550/arXiv.1207.3907
Hallsdóttir M. On the pre-settlement history of Icelandic vegetation. *Búvísindi (Icel Agric Sci)*. 1995;9:17–29.
Hamrick JL, Godt MJW. Effects of life history traits on genetic diversity in plant species. *Philos Trans R Soc Lond B Biol Sci*. 1996;351:1291–1298.
Hill JL, Hollender CA. Branching out: new insights into the genetic regulation of shoot architecture in trees. *Curr Opin Plant Biol*. 2019;47:73–80.
Hu Y, Chen X, Zhou T, Hou N, Zhao P. Characterization of the complete chloroplast genome of an endangered species dwarf birch (*Betula nana* L.). *Conserv Genet Resour*. 2017;9:55–58.
Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet*. 2010;11:94.
Jónsson TH. Stature of sub-arctic birch in relation to growth rate, lifespan and tree form. *Ann Bot*. 2004;94:753–762.
Knaus BJ, Grünwald NJ. VCFR: a package to manipulate and visualize variant call format data in R. *Mol Ecol Res*. 2017;17:44–53.
Kosman E, Leonard KJ. Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploidy species. *Mol Ecol*. 2005;14:415–424.
Kremer A. Genetic diversity and phenotypic variability of forest trees. *Genet Sel Evol*. 1994;26:S105–S123.
Kremer A, Ronce O, Robledo-Arnuncio JJ, Guillaume F, Bohrer G, Nathan R, Bridle JR, Gomulkiewicz R, Klein EK, Ritland K, et al. Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecol Lett*. 2012;15:378–392.
Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10:R25.
Meucci S, Schulte L, Zimmermann HH, Stoof-Leichsenring KR, Epp L, Bronken Eidesen P, Herzschuh U. Holocene chloroplast genetic variation of shrubs (*Alnus alnobetula*, *Betula nana*, *Salix* sp.) at the Siberian tundra-taiga ecotone inferred from modern chloroplast genome assembly and sedimentary ancient DNA analyses. *Ecol Evol*. 2021;11:2173–2193.
Oksanen E. Birch as a model species for the acclimation and adaptation of northern forest ecosystem to changing environment. *Front For Glob Change*. 2021;4:682512.
Oksanen J, Blanchet FG, Kindt R, et al. 2014. Vegan: community ecology ackage. R Package Version 2.2-0. http://CRAN.Rproject.org/package=vegan
Ottósson JG, Sveinsdóttir A, Harðardóttir M, editors. *Vistgerðir á Íslandi. Fjölrit Náttúrufræðistofnunar nr. 54*. 2016. p. 1–299. Garðabær: Icelandic Institute of Natural History. https://utgafa.ni.is/fjolrit/Fjolrit_54.pdf
Palmé AE, Su Q, Palsson S, Lascoux M. Extensive sharing of chloroplast haplotypes among European birches indicates hybridization among *Betula pendula*, *B. pubescens* and *B. nana*. *Mol Ecol*. 2004;13:167–178.
Paradis E. *pegas*: an R package for population genetics with an integrated-modular approach. *Bioinformatics*. 2010;26:419–420.
Przedpelska-Wasowicz EM, Wasowicz P, Áskelsdóttir AO, Guðjohnsen ER, Hallsdóttir M. Characterization of pollen seasons in Iceland based on long-term observations: 1988–2018. *Aerobiologia*. 2021;37:507–524.
R Core Team. *R: a language and environment for statistical computing*. Vienna (Austria): R Foundation for Statistical Computing; 2020. https://www.R-project.org/
Razgour O, Forester B, Taggart JB, Bekaert M, Juste J, Ibáñez C, Puechmaille SJ, Novella-Fernandez R, Alberdi A, Manel S. Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. *Proc Natl Acad Sci USA*. 2019;116:10418–10423.
Salojärvi J, Smolander OP, Nieminen K, Rajaraman S, Safronov O, Safdari P, Lamminmäki A, Immanen J, Lan T, Tanskanen J, et al. Genome sequencing and population genomic analyses provide insights into the adaptive landscape of silver birch. *Nat Genet*. 2017;49:904–912.
Schlaepfer DR, Braschler B, Rusterholz HP, Baur B. Genetic effects of anthropogenic habitat fragmentation on remnant animal and plant populations: a meta-analysis. *Ecosphere*. 2018;9(10):e02488.
Snorrason A, Traustason B, Kjartansson, BÞ, Heiðarsson L, Ísleifsson R, Eggertsson Ó. Náttúrulegt birki á Íslandi—ný úttekt á útbreiðslu þess og ástandi. *Náttúrufræðingurinn*. 2016;86:97–111.
Taylor JJ, Lawler JP, Aronsson M, Barry T, Bjorkman AD, Christensen T, Coulson SJ, Cuyler C, Ehrich D, Falk K, et al. Arctic terrestrial biodiversity status and trends: a synopsis of science supporting the CBMP State of Arctic Terrestrial Biodiversity Report. *Ambio*. 2020;49:833–847.
Thomson A, Dick C, Pascoini A, Dayanandan S. Despite introgressive hybridization, North American birches (*Betula* spp.) maintain strong differentiation at nuclear microsatellite loci. *Tree Genet Genomes*. 2015;11:1–12.
Thórsson ÆTh, Pálsson S, Lascoux M, Anamthawat-Jónsson K. Introgression and phylogeography of *Betula nana* (diploid), *B. pubescens* (tetraploid) and their triploid hybrids in Iceland inferred from cpDNA haplotype variation. *J Biogeogr*. 2010;37:2098–2110.
Thorsson ÆT, Pálsson S, Sigurgeirsson A, Anamthawat-Jónsson K. Morphological variation among *Betula nana* (diploid), *B. pubescens* (tetraploid) and their triploid hybrids in Iceland. *Ann Bot*. 2007;99:1183–1193.
Tomczyk P, Kolodziejczyk I, Heddicke J, Ólafsson E, Wasowicz P. An attempt to quantify leaf damage in *Betula pubescens* by non-native leaf miners: *Heringocrania unimaculella* and *Scolioneura betuleti* in Akureyri (N Iceland). Poster presented at IceBio 2017 conference; 2017.
Tsuda Y, Semerikov V, Sebastiani F, Vendramin GG, Lascoux M. Multispecies genetic structure and hybridization in the *Betula* genus across Eurasia. *Mol Ecol*. 2017;26:589–605.
Turner IM. Species loss in fragments of tropical rain forest: a review of the evidence. *J Appl Ecol*. 1996;33:200–209.
Walters SM. Betula. 1964. In: Tutin TG, et al., editors. *Flora Europaea*. Vol. 1. Lycopodiaceae to Platanaceae, p. 464. Cambridge University Press. doi:10.5281/zenodo.302862
Wang N, Kelly LJ, McAllister HA, Zohren J, Buggs RJA. Resolving phylogeny and polyploid parentage using genus-wide genome-wide sequence data from birch trees. *Mol Phylogenet Evol*. 2021;160:107126.
Wasowicz P. Annotated checklist of vascular plants of Iceland. *Fjölrit Náttúrufræðistofnunar*. 2020;57:1–193.
Wasowicz P, Sennikov AN, Westergaard KB, Spellman K, Carlson M, Gillespie LJ, Väre H. Non-native vascular flora of the arctic: taxonomic richness, distribution and pathways. *Ambio*. 2020;49:693–703.
Yang X-Y, Wang Z-F, Luo W-C, Guo X-Y, Zhang C-H, Liu J-Q, Ren GP. Plastomes of Betulaceae and phylogenetic implications. *J Syst Evol*. 2019;57:508–518.
Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. *J Comput Biol*. 2000;7:203–214.
Zohren J, Wang N, Kardailsky I, Borrell JS, Joecker A, Nichols RA, Buggs N. Unidirectional diploid–tetraploid introgression among British birch trees with shifting ranges shown by restriction site-associated markers. *Mol Ecol*. 2016;25:2413–2426.