Comparative analyses of plastid and AFLP data suggest different colonization history and asymmetric hybridization between Betula pubescens and B. nana

Article *in* Molecular Ecology · June 2015

DOI: 10.1111/mec.13289 · Source: PubMed

CITATIONS
9

READS
282

3 authors, including:

Pernille Bronken Eidesen
University Centre in Svalbard (UNIS)
48 PUBLICATIONS **2,123** CITATIONS

SEE PROFILE

Inger Greve Alsos
UiT The Arctic University of Norway
161 PUBLICATIONS **2,721** CITATIONS

SEE PROFILE

Received Date : 16-Mar-2015

Revised Date : 17-Jun-2015

Accepted Date : 18-Jun-2015

Article type : Original Article

Title:

Comparative analyses of plastid and AFLP data suggest different colonization history and asymmetric hybridisation between *Betula pubescens* and *B. nana*.

Pernille Bronken Eidesen[1], Inger Greve Alsos[2], Christian Brochmann[3]

[1]The University Centre in Svalbard, PO Box 156, NO-9171 Longyearbyen, Norway

[2]Tromsø University Museum, NO-9037 Tromsø, Norway

[3]Natural History Museum, University of Oslo, PO Box 1172 Blindern, NO-0318 Oslo, Norway

Correspondence: Pernille Bronken Eidesen

The University Centre in Svalbard, PO Box 156, NO-9171 Longyearbyen, Norway

Tlf: +47 79 02 33 43

Fax: +47 79 02 33 01

E-mail: pernillee@unis.no

Short running head: Postglacial colonization of *Betula*

Keywords: AFLP, plastid DNA, introgression, glacial refugia, migration, phylogeography

Abstract

Birches (*Betula* spp.) hybridise readily, confounding genetic signatures of refugial isolation and postglacial migration. We aimed to distinguish hybridisation from range-shift processes in the two widespread and cold-adapted species *B. nana* and *B. pubescens*, previously shown to share a similarly east-west-structured variation in plastid DNA (pDNA). We sampled the two species throughout their ranges and included reference samples of five other *Betula* species and putative hybrids. We analysed 901 individual plants using mainly nuclear high-resolution markers (amplified fragment length polymorphisms; AFLPs); a subset of 64 plants was also sequenced for two pDNA regions. Whereas the pDNA variation as expected was largely shared between *B. nana* and *B. pubescens*, the two species were distinctly differentiated at AFLP loci. In *B. nana*, both the AFLP and pDNA results corroborated the former pDNA-based hypothesis that it expanded from at least two major refugia in Eurasia, one south of and one east of the North European ice sheets. In contrast, *B. pubescens* showed a striking lack of geographic structuring of its AFLP variation. We identified a weak but significant increase in nuclear (AFLP) gene flow from *B. nana* into *B. pubescens* with increasing latitude, suggesting hybridisation has been most frequent at the

postglacial expansion front of *B. pubescens* and that hybrids mainly backcrossed to *B. pubescens*. Incongruence between pDNA and AFLP variation in *B. pubescens* can be explained by efficient expansion from a single large refugium combined with leading-edge hybridisation and plastid capture from *B. nana* during colonization of new territory already occupied by this more cold-tolerant species.

Introduction

Birches (*Betula* L.) provide excellent examples of how hybridisation may confound both species boundaries and phylogeographic patterns (Palmé *et al.* 2004; Anamthawat-Jónsson 2012; Wang *et al.* 2014; Thomson *et al.* 2015). Plastid DNA (pDNA) haplotypes are typically extensively shared among *Betula* species. Although shared haplotypes may result from recurring mutations or incomplete lineage sorting, the available evidence strongly supports hybridisation and introgression as the main explanation for haplotype sharing in *Betula* (Palmé *et al.* 2004). The taxonomy of the genus is notoriously difficult, partly due to frequent hybridisation (Jonsell & Karlsson 2000; Thórsson *et al.* 2001; Järvinen *et al.* 2004; Schenk *et al.* 2008; Ashburner & McAllister 2013), and phylogenetic analyses have shown incongruence between plastid and nuclear DNA sequence variation (Järvinen *et al.* 2004). Amplified Fragment Length Polymorphisms (AFLPs) have also been used to infer phylogenetic relationships in *Betula*. These markers revealed similar relationships as nuclear DNA sequences, but suggested that the two cold-tolerant species *B. nana* and *B. pubescens* are more closely related than inferred from former studies (Schenk *et al.* 2008).

Four species of *Betula* are usually recognised in Europe, and are capable of hybridising with each other. Two are diploid shrubs, the eastern European and Siberian *B. humilis* Schrank. and the circumpolar *B. nana* L., while two are Eurasian trees, the diploid *B. pendula* Roth and the

tetraploid *B. pubescens* Ehrh. (Tutin 1964; Atkinson 1992; DeGroot *et al.* 1997; Palmé *et al.* 2004). *Betula pendula* and *B. pubescens* are nearly sympatric, but *B. pubescens* is more cold-tolerant and coexists with shrub birches at high latitudes and altitudes. At high latitudes, *B. pubescens* typically co-occurs with *B. nana*, which is the hardiest of the European birches and common in low- arctic regions; in contrast, *B. humilis* does not reach the Arctic (Elven *et al.* 2011). All four species have earlier been subjected to phylogeographic analyses based on pDNA markers (Palmé 2003; Palmé *et al.* 2004; Maliouchenko *et al.* 2007; Thórsson *et al.* 2010; Jadwiszczak *et al.* 2012; Jadwiszczak *et al.* 2014). Two major haplotypes were found throughout Europe in all species, one dominating in western Europe and one in eastern Europe. This pattern suggests two main waves of postglacial recolonization; both from intermediate latitudes, but from different directions (Palmé 2003; Jadwiszczak 2012). However, the extensive pDNA haplotype sharing inferred to be due to hybridisation, in particular among *B. pendula*, *B. pubescens*, and *B. nana* (Palmé *et al.* 2004), may mislead interpretation of their species-specific phylogeographies (Jadwiszczak 2012).

Plastid DNA introgresses more frequently and faster than nuclear genes, and it can introgress without significant nuclear gene flow (i.e. chloroplast capture; Rieseberg & Soltis 1991; Rieseberg *et al.* 1996). A recent study of the postglacial history of *B. pendula*, *B. pubescens*, and *B. nana* in the British Isles used nuclear microsatellites instead of pDNA markers and showed the three species to be clearly distinguished with only low levels of introgression evident between them (Wang *et al.* 2014). Nuclear introgression was detected from *B. nana* into *B. pubescens*, and increased with increasing latitude in the study area. As simulations suggest that introgression occurs almost exclusively from a local to an invading species (Currat *et al.* 2008), the introgression pattern in British *B. pubescens* most likely results from past northwards range expansion of *B. pubescens* into areas already occupied by the more cold-adapted *B. nana*. In addition, the southern range-contraction of *B. nana* is likely to have further reduced introgression at lower

latitudes. Wang *et al*. (2014) argued that alternative explanations to introgression, such as convergent mutations and incomplete lineage sorting, were unlikely. The asymmetric allele sharing rather supported introgression than convergent mutations, and the gradual increase in number of shared alleles closer to the current location of *B. nana* populations supported introgression rather than incomplete lineage sorting.

At a larger geographical scale, it remains unclear to what degree these two birch species have similar phylogeographic histories and to what degree one of them has acquired the plastids of the other when expanding into its area. Studies of animals and plants based on both organelle DNA and nuclear markers demonstrate that erroneous conclusions can be drawn from a single marker if historical hybridisation has occurred (e.g. Shaw 2002; Eidesen *et al.* 2007a), and Palmé *et al*. (2004) and Anamthawat-Jónsson (2012) have called for analyses of nuclear markers to further explore species' phylogeographies in *Betula*.

In the present study, we aimed to disentangle the glacial and postglacial history of the hardiest birches in Europe, *B. nana* and *B. pubescens*, in terms of refugial isolation, migration and hybridisation. We sampled *B. nana* and *B. pubescens* throughout most of their geographic ranges (Fig. 1), and also included available samples of four other birch species and some putative hybrids, so as to help in the identification of introgression (Rieseberg & Soltis 1991). We complemented plastid DNA analysis with analysis of genomic DNA by AFLPs (Vos *et al.* 1995), which mainly represent nuclear DNA, and compared our plastid- and AFLP results with the previously published plastid DNA-based patterns. Based on the extensive evidence from former pDNA studies, which discard recurrent mutations or incomplete lineage sorting as explanations for shared plastid haplotypes in *Betula* (e.g. Palmé *et al.* 2004; Thomson *et al.* 2015), we here assume that plastid

haplotypes shared among species result from hybridisation events. Moreover, based on the results of the recent nuclear analysis of microsatellite variation in British *Betula* (Wang *et al.* 2014), we further hypothesise that AFLPs will be less influenced by introgression than pDNA markers. Thus, we expect that AFLP data should distinguish the species and that geographical structuring of AFLP variation within species will largely reflect refugial isolation and postglacial migration history rather than hybridisation events. If plastid patterns of variation were also influenced by refugial isolation and post-glacial migration history, we would expect AFLP and plastid data to show congruent patterns. Alternatively, if there is a discrepancy between AFLP and plastid patterns, this would imply that the plastid pattern of variation has been influenced by hybridisation events such that it cannot be directly interpreted in terms of refugial isolation and postglacial migration.

Material and Methods

Study species

The genus *Betula* (Betulaceae) consists of monoecious, wind-pollinated, and wind-dispersed trees and shrubs. *Betula nana* ($2n = 2x = 28$) is a dwarf shrub native to arctic and boreal regions with a circumpolar distribution (Fig. 1). It typically occurs in moist, acidic, and nutrient-poor sites (DeGroot *et al.* 1997), and at southern latitudes is usually found only in mountain areas (Hultén & Fries 1986). Two largely allopatric subspecies are usually recognised: ssp. *nana* (Suk.) Hult. in Greenland, Europe and western Asia, and ssp. *exilis* (Suk.) Hult. in North America and central and eastern Asia (Hultén & Fries 1986; Elven *et al.* 2011). In this study we have followed the interpretation by Elven *et al*. (2011), but it should be noted that *B*. *nana ssp. exilis* is also referred to as *B. glandulosa* according to the latest classification of *Betula* (Ashburner & McAllister, 2013).

Betula pubescens (2*n* = 4*x* = 56) is a Eurasian birch species (Fig. 1; Elven *et al.* 2011), with wide climatic tolerance (Atkinson 1992). In Europe, it is usually divided into two subspecies: ssp. *pubescens*, found in the lowlands with the same distribution as *B. pendula*, and ssp. *tortuosa* (Ledeb.) Nyman, restricted to upland areas and overlapping with *B. nana* (Elven *et al.* 2011). In the current study, most populations of *B. pubescens* were sampled from upland areas or at high latitudes in Eurasia and belong to ssp. *tortuosa*; thus, we consider hybridisation to be more likely with *B. nana* than with the lowland species *B. pendula*.

For use as reference species, we included samples available of *B. pendula* Roth (2*n* = 2*x* = 28), and three North American species, *B. pumila* L. (2*n* = 4*x* = 56; a North American shrub), *B. michauxii* Spach. (a south-east Canadian dwarf shrub) and *B. glandulosa* Michx. (2*n* = 2*x* = 28; a North American shrub). In addition, we included from Greenland, samples of what had been classified as *B*. cf. *minor* (2*n* = 4*x* = 56), which is regarded as hybrid between *B. glandulosa* x *B. cordifolia* (Ashburner & McAllister, 2013). *Betula glandulosa* is closely related to *B. nana* and has similar habitat preferences, but it is somewhat less cold tolerant (DeGroot *et al.* 1997). *Betula glandulosa* and *B. nana* are known to hybridise, sometimes resulting in a continuum of intermediate forms where their ranges overlap (Fig. 1; (Furlow 1997).

AFLP datasets

We re-scored and re-analysed two available AFLP datasets, one mainly containing populations of *B. nana* s.l. (570 individuals), the other mainly containing populations of *B. pubescens* (332 individuals; Table S1, Supporting Information). These datasets were initially generated and analysed as part of several larger meta-studies (Alsos *et al.* 2007; Alsos *et al.* 2012; Eidesen *et al.* 2013; Alsos *et al.* 2015). The *B. pubescens* dataset also, included populations of other *Betula*

species omitted from previously published analyses; i.e. *B.* cf. *major*, *B. pendula*, *B. pumila*, *B. glandulosa*, and *B. michauxii*, and some putative hybrids as inferred from morphology (*B.* cf. *pendula* x *pubescens* and *B.* cf. *nana* x *pubescens*).

Sample collection for these datasets followed a standardized scheme, where leaves from up to 11 plants per species were collected 25 m apart along a 250 m transect per sample location. As 25 m is less than estimates of the seed shadow of *Betula*, some bias due to spatial autocorrelation might be present. Details regarding sampling schemes, DNA extraction and AFLP procedures are provided in the online Supporting Information.

Each of the two original datasets was initially constructed using three AFLP primer combinations (after testing 40-53 combinations). Both datasets had one primer combination in common (6-FAM labelled *Eco*R1-AGT/*Mse*1-CTC), and for this primer combination, we imported the original AFLP profiles from both datasets into Genographer, version 1.6., re-scored all samples, and combined them into one dataset. Unambiguous, polymorphic fragments in the size range of 50–500 base pairs were scored as present (1) or absent (0). This new dataset, called the "*Betula* spp. dataset", included 901 samples from 117 populations, covering the entire geographic ranges of *B. nana* and *B. pubescens* s. lat. (Fig. 1, Table S1). We scored 104 polymorphic markers, with an error rate of 1.7% calculated from 81 replicates (Bonin *et al.* 2004). The *Betula* spp. dataset contained far more individuals than the original datasets, but fewer markers, which could potentially reduce the power of the dataset. Mantel tests, Principal Coordinate analyses (Fig. S1), and Neighbour Joining analyses (Figs. S2 and S3) were used to evaluate consistency among the two original datasets and the re-scored dataset (Supporting Information). Although the power was reduced, the new *Betula* spp. dataset revealed genetic patterns concordant with those of the

two original datasets and was regarded reliable for further analyses (Supporting Information).

To ensure maximum resolution within our focal species in the present study, *B. nana* and *B. pubescens*, we also analysed these species separately based on subsets of the original datasets generated with all three primer combinations. After excluding the species reference samples and hybrids, our *B. nana* dataset included 528 individuals scored for 115 polymorphic markers, and our *B. pubescens* dataset included 303 individuals scored for 132 polymorphic markers (Table S1, Supporting Information). Error rates were 2.0% (51 replicates) and 0.6% (30 replicates), respectively.

Thus, further data analyses were based on three different AFLP matrices: 1) the *Betula* spp. dataset (all samples from the two original datasets combined and re-scored for one primer combination), 2) the *B. nana* dataset (a subset of the original *B. nana* dataset based on three primer combinations), and 3) the *B. pubescens* dataset (a subset of the original *B. pubescens* dataset based on three primer combinations).

AFLP data analyses

If not specified otherwise, data analyses were performed on all three AFLP datasets constructed for this study as outlined above. Genetic groups were inferred using STRUCTURE version 2.3.4 (Pritchard *et al.* 2000), where "0" was indicated as the recessive allele in the binary AFLP matrix (Falush *et al.* 2007). As our focal species have wide distributions, are pollinated and dispersed by wind, and tend to hybridise, the appropriate model should handle closely related groups and assume both recent ancestry and current gene flow. Thus, we used the admixture model with correlated allele frequencies (Falush *et al.* 2003), and ran the analyses with and without the

LOCPRIOR option, which utilizes the sampling location for clustering, and can enhance structuring of weak groups (Hubisz *et al.* 2009). A burn-in period of 2×10^5 iterations was used followed by 10^6 iterations. Ten replicates were run for each *K* from 1 to 10, and evaluation of number of groups was done according to recommendations given by Pritchard *et al*. (2000) and Evanno *et al.* (2005). Summaries of output were visualised by the online software STRUCTURE HARVESTER (Earl & vonHoldt, 2012). Genetic groups inferred by STRUCTURE, were also verified by another Bayesian clustering algorithm run in BAPS Version 6.0 (Corander *et al.* 2003).

Spatial distribution of genetic groups was inspected visually in ArcGIS 10, and the relationship between level of admixture (STRUCTURE values) between species (*B. nana* and *B. pubescens*) and latitude was tested as described in Wang *et al.* (2014), using a mixed effects model (lme) in R v. 3.1.0 (R Core Team 2014), with slope as a fixed effect, and population modelled as a random effect to allow for genetic drift. STRUCTURE values were logit-transformed to improve residual fit.

A Neighbour joining (NJ) analysis in PAST v. 3.0 (Hammer *et al.* 2001) was conducted on the *Betula* spp. dataset, using populations as OTUs and the population frequencies of each AFLP marker as variables (Eidesen *et al.* 2007b). Although populations do not evolve in a bi-furcating manner, this clustering approach reduces the complexity of datasets comprising many individuals, and examines similarity within and between species. Populations with less than three individuals were omitted from this analysis. Marker frequencies were calculated using the R-script package AFLPdat (Ehrich 2006), *Betula pumila* was used as outgroup, similarity index was set to Bray-Curtis, and bootstrap resampling was done with 1000 replicates.

Principal coordinates analyses (PCO) executed in PAST with Jaccard similarity index were used to visualize pairwise similarity among the AFLP multilocus phenotypes.

To estimate differentiation between populations and groups, AFLP multilocus phenotypes were treated as haplotypes to create a phenotypic (band-based) distance matrix. Pairwise F_{st} and analyses of molecular variance (AMOVA; Excoffier *et al.* 1992) based on pairwise distances (Simple matching) were run in ARLEQUIN v. 3.5.1.3 (Excoffier & Lischer, 2010).

Calculation of intra-population diversity (Nei 1987; Kosman 2003), and a rarity index (a measure of the amount of rare markers in each population (Schönswetter & Tribsch 2005; Ehrich 2006) were calculated by R-scripts provided in AFLPdat. To reduce the influence of rare markers resulting from hybridization, and to emphasise historical processes when inferring phylogeographical patterns within species, we based our analyses of rarity patterns on the results from the *Betula* spp. dataset only.

To compare means of diversity and rarity among species and genetic groups, analyses of variance (ANOVA) were performed in PAST, followed by Tukey's pairwise post-hoc tests. Spatial patterns of diversity and rarity were inspected visually by plotting values geographically in ArcGIS 10. Possible relationships between genetic diversity and/or rarity within species and latitude and/or longitude (adjusted to vary as a continuous positive variable from east to west throughout the distribution range) were investigated by multiple linear regression analyses in PAST.

Correlations between genetic and geographical distances were computed for the *B. nana* and *B. pubescens* datasets. Isolation by distance was tested by Mantel tests (Mantel 1967) in PAST. For each dataset, geographical distances between populations were recalculated from geographical coordinates using Geographic Distance Matrix Generator v. 1.2.3 (Ersts), and tested against the corresponding pairwise F_{st} values retrieved from ARLEQUIN.

Sequencing and analyses of plastid DNA

A subset of 64 individuals from 59 populations was chosen for sequencing based on the genetic structure revealed by the AFLP analyses (Table S1, Supporting Information). We initially screened eight universal pDNA primer pairs: c-f (Taberlet *et al.* 1991), psbA-*trn*H, *trn*S-*trn*G (Hamilton 1999), rpoB-*trn*C, *trn*C-ycf6_R, ycf6_F-*trn*D (Shaw *et al.* 2005), rpl32-*trn*L and *trn*Q- rsp16 (Shaw *et al.* 2007). Based on amplification efficiency and level of variation, two pDNA regions, *trn*Q-*rps*16 and *trn*S-*trn*G, were chosen for further sequencing. A large indel area in *trn*S-*trn*G pDNA was avoided by exchanging the *trn*G primer with another reverse primer, *trn*BW (5`TTAG AGA TTG ACG TGC TTT GTT 3`) (Alsos 2003). Samples were amplified by PCR using 10x diluted DNA template following standard procedures and amplification parameters from Shaw *et al.* (2007). PCR products were purified using ExoSAP-IT Kit (Amersham Biosciences), and sequenced in both directions using Applied Biosystems BigDye Terminator v.1.1 Kit. Sephadex G-50 (Amersham Biosciences) was used to remove unincorporated dye terminators prior to analysis on an ABI 3100 Genetic Analyzer. DNA sequences were edited and aligned in Sequencher 4.1.4 (Genecodes, Ann Arbor, Michigan, USA). Sequences have been deposited in GenBank under accession nos. KR779549 - KR779606 and KR779607 - KR779658. Most variation was due to indels, including nested indels. We scored variable sites as binary characters (substitutions and indels) or unordered, multi-state characters when gaps and substitutions were found in the same position.

Ambiguous poly-T regions were omitted. Plastid haplotypes were defined by combining information from the two sequenced regions. Geographical distributions of haplotypes were inspected visually in ArcGIS 10.

Relationships among pDNA haplotypes were assessed using Neighbour joining analysis in PAST using Hamming distance, *B. pumila* as outgroup, and bootstrap resampling with 1000 replicates. As phylogenetic methods may reveal limited resolution at the intraspecific level (Posada & Crandall 2001), the relationships among pDNA haplotypes were also assessed by constructing a median-joining network using the software NETWORK v. 4.6.1.2 (Bandelt *et al.* 1999), which allows multi-state data to be included. The matrix was entered as mixed data, and run using default settings.

Correlations between the AFLP and pDNA datasets were computed using Mantel tests in PAST, where genetic distances among AFLP multilocus phenotypes based on Jaccard similarity were tested against the corresponding genetic distances among pDNA haplotypes based on Hamming distance. Mantel tests were also used to test possible correlations among geographical and genetic distances based on pDNA variation.

Associations among haplotype frequencies and species, and among haplotype frequencies and genetic groups, were investigated by contingency tables in PAST, with p values based on Fisher's exact test (two-tailed).

Results

The *Betula* spp. AFLP dataset

Similar main partitions were inferred based on BAPS- and STRUCTURE analyses, run both with and without locality information included. Thus, only data from STRUCTURE analyses with locality information are presented. Four genetic groups were inferred based on where lnP(Data) plateaued, and delta K clearly peaked (Fig. 2a; Fig. S4a, Supporting Information). These four groups corresponded well to the four dominant species/subspecies studied and their respective geographical distributions: "*B. glandulosa*", "*B. pubescens*", "*B. nana* East" (largely corresponding to the American/East Siberian ssp. *exilis*), and "*B. nana* West" (largely corresponding to the Atlantic/West Siberian ssp. *nana*) (Fig. 1; Fig. S5a-d, Table S2, Supporting Information). The two *B. nana* groups showed high levels of admixture, particularly in West Siberia where the ranges of the two subspecies overlap (Figs. 1 and 2 a). The same clustering into species and subspecies were reflected in the NJ tree, but with low support (Fig. S3, Supporting Information). Reference species only represented by one population were not recognized as separate genetic groups by STRUCTURE (Table S2, Supporting Information), but were placed however outside the four main clusters in the NJ tree (Fig. S3, Supporting Information).

The *B. glandulosa* group was the most divergent of the four groups (PCO and heat map; Fig. S5e-f, Supporting Information), and hierarchical AMOVA analysis conducted after combining *B. nana* samples into one group and excluding hybrids and reference species represented by a only a single population showed that 22.3% of the total variation was due to differences between *B. glandulosa*, *B. pubescens* and *B. nana* and 13% resulted from differences among populations. Mean population diversity was similar in these three groups (0.084, SD ± 0.044, 0.088 ± 0.019, and 0.085 ± 0.016, respectively), while rarity was considerably higher in *B. glandulosa* (0.227 ± 0.093 versus 0.096 ± 0.040 and 0.111 ± 0.050).

The level of admixture between our focal species was low overall, but varied among populations and species. On average, twice as much admixture was inferred to have occurred into *B. pubescens* from *B. nana* than vice versa (0.053 ± 0.055 versus 0.026 ± 0.037; t-test; $p < 0.01$; Fig. 2, Table S2, Supporting Information), and the level of admixture showed a significant and positive linear correlation with latitude (Fig. 3a; $p < 0.001$). There was no trend related to latitude in *B. nana* (Fig. 3b; $p = 0.999$).

The *Betula nana* AFLP dataset

Five genetic groups were inferred to be present in *B. nana* from the STRUCTURE analyses of the *B. nana* dataset (Fig. 2 b; Fig. S4b, Supplementary Information). Although Delta K showed no clear peak, BAPS showed the highest likelihood for five groups with the same partition that was indicated by STRUCTURE. The five groups showed a clear east-west geographical pattern: 1) Beringia - East Siberia, 2) Svalbard, 3) West Siberia, 4) Europe, and 5) the Atlantic area (Fig. 4 a). PCO showed a similar east-west geographical pattern along axis 1 (8.2 % of the variation, Fig. S6. Supporting Information). In the hierarchical AMOVA, 13.0% of the total variation was found among the five groups and 10.0% was found among populations within groups (group affinity based on average assignment value for each population; Table S2, Supporting Information). Overall population differentiation (F_{st}) was 0.205, and pairwise F_{st} values between populations were correlated with geographical distances (Mantel test, $r = 0.479$, $p < 0.001$).

Because of the increased number of polymorphic markers for *B. nana* in this dataset compared to the *Betula* spp. dataset, it showed higher intrapopulation diversity (mean over all populations 0.138, ± 0.025), but the values from the two datasets were strongly correlated ($r = 0.752$, $p < 0.001$). The mean intrapopulation diversities in the five genetic groups were

significantly different (ANOVA; $p < 0.05$) because of the low diversity in the Svalbard group (0.106 ± 0.015; Tukey's pairwise post-hoc tests) compared to the others (Atlantic: 0.127 ± 0.039, Europe: 0.142 ± 0.020, West Siberia: 0.151 ± 0.018, and Beringia – East Siberia: 0.146 ± 0.014). We found a strong positive linear relationship between latitude and diversity for populations assigned to the European group (slope: 0.0021, intercept: 0.015; $r = 0.672$, $p < 0.001$), whereas the Atlantic group showed the opposite trend (slope: -0.0062, intercept: 0.5513; $r = -0.548$, $p = 0.05$; Fig. 4a).

No linear relationships with latitude or longitude were found by regression analyses (results not shown). Although highest average rarity was found in the mainly unglaciated Beringia - East Siberia group (0.127 ± 0.034), followed by West Siberia (0.114 ± 0.032), Europe (0.088 ± 0.043), Atlantic (0.087 ± 0.037) and Svalbard (0.086 ± 0.030), the difference were not significant (ANOVA; $p = 0.08$; Fig. S7a, Table S2, Supporting Information).

The *Betula pubescens* AFLP dataset

Even with the higher number of markers in this dataset compared to the *Betula* spp. dataset, there was a striking absence of genetic structure in *B. pubescens.* Neither BAPS nor STRUCTURE analyses suggested any partition of the dataset (Fig. 4b). The populations were only weakly differentiated as inferred from F_{st} (0.058), and genetic and geographical distances were weakly correlated (Mantel test, $r = 0.242$, $p < 0.001$). Adding more markers increased the intrapopulation diversity levels (mean over all populations 0.157, ± 0.025; Table S2, Supporting Information), but they were strongly correlated with the values based on the *Betula* spp. dataset ($r = 0.856$, $p < 0.001$). Intrapopulation diversity (Fig. 5) and rarity (Fig. S7b, Supporting Information) showed no clear geographical structure, or linear relationships with latitude or longitude as determined by regression analyses (results not shown).

The plastid DNA dataset

The *trn*Q-*rps*16 region was successfully amplified for 54 samples and was 640 bp in length after trimming and alignment. The *trn*-S-*trn*G region was successfully amplified for 58 samples and was 502 bp in length after trimming and alignment. Eight and ten variable sites were identified in the two pDNA regions, of which most were length variations (13 indels, 2 substitutions, and 3 sites with gaps and substitution in the same position). Both regions were successfully amplified for 48 samples, and 14 pDNA haplotypes were identified (Table S3, Supporting Information).

In the haplotype network and NJ analysis, the pDNA haplotypes with the *trn*-S-*trn*G variant G (Ge, Gh, Gg) were strongly differentiated from the rest (Fig. 6; Fig. S8, Supporting Information). These haplotypes were found in the Canadian *B. pumila* and in the Greenlandic member of the *B. pubescens* aggregate, *B.* cf. minor (Table 1). The Canadian *Betula michauxii* also had the *trn*S-*trn*G variant G, but was not successfully analysed for *trn*Q-*rps*16. In the two other reference species, *B. pendula* (not successfully analysed for *trn*Q-*rps*16) and *B. glandulosa*, only haplotypes with the frequent *trn*-S-*trn*G variant A were found (Table 1). Haplotype Aa was most frequent and geographically widespread of the haplotypes and placed as central in the network (Figs 4 c, d and 6). This haplotype was shared among *B. glandulosa*, *B. nana* and *B. pubescens*. It dominated in the Beringian - East Siberian area, but was also frequent in Europe. The other haplotypes were more geographically restricted. Haplotype Cc and Ca were shared between *B. nana* and *B. pubescens* and dominated in West Siberia – Eastern Europe and also occurred in Central Europe. Ab was only found in *B. nana*, and completely dominated Atlantic areas like Greenland where *B. pubescens* is not present. In our focal species, *B. nana* and *B. pubescens*, the highest haplotype diversity was found in Europe, and three haplotypes were restricted to the British Isles (Fig. 4c, d).

There was no correlation between genetic distances based on AFLP and pDNA, or between the geographical and genetic distances between the pDNA haplotypes. The test of contingency tables showed a significant difference in distribution of pDNA haplotypes between *B. nana* and *B. pubescens* (Fisher's exact test $p < 0.01$), but if the *B. nana* specific Atlantic haplotype Ab was left out of the analyses, the differences were no longer significant. In *B. nana*, the geographic distribution of different pDNA haplotypes was associated with different AFLP groups (Fisher's exact test $p < 0.001$); the Beringia–East Siberia group was completely dominated by the assumed ancient haplotype Aa, the West Siberia and Svalbard groups were dominated by the tip haplotype Cc, while the Atlantic group was completely dominated by the tip haplotype Ab. In Europe, all haplotypes were present, but the assumed ancient haplotype Aa was most common.

Discussion

In this study, we were able to disentangle the glacial and postglacial history of *B. nana* and *B. pubescens* in terms of refugial isolation, migration and hybridisation by using range-wide sampling and a combination of plastid and nuclear DNA markers. We have demonstrated that there is a striking incongruence between pDNA and AFLP variation in *B. pubescens* whereas its close, but more cold-tolerant relative *B. nana* shows congruent geographic structuring of its organellar and nuclear variation. Furthermore, the two species are clearly differentiated at nuclear loci but not to the same extent in their plastid DNA. We argue here that the two species have different phylogeographic histories, that the parallel geographic structuring of the shared pDNA variation in these two species results from postglacial hybridisation between them, and that the increase we found in nuclear admixture from *B. nana* into *B. pubescens* with increasing latitude suggests that hybridisation has facilitated a northwards postglacial expansion of *B. pubescens* into territory already occupied by the more cold-tolerant *B. nana*.

Our data confirm former reports of pDNA haplotype sharing and extensive hybridisation among different species of *Betula* (Jadwiszczak 2012; Thomson *et al.* 2015). Three common pDNA haplotypes were shared among species and these haplotypes showed similar geographical distributions in *B. nana* and *B. pubescens* (Fig. 4 c, d; Table 1). Besides hybridisation, shared pDNA among closely related species can result from incomplete lineage sorting of ancestral polymorphisms or from parallel mutations. As plastid DNA evolves slowly, parallel mutations have been discarded as a likely explanation for the extensive haplotype sharing in *Betula* (Palmé *et al.* 2004). Our haplotype Aa was widespread and common, inferred as central in the network, and also found in *B. glandulosa*, which was clearly divergent based on AFLPs. Thus, this haplotype may represent shared, ancient variation. Common tip haplotypes such as Cc, on the other hand, which has a similar geographical distribution in *B. nana* and *B. pubescens*, are much less likely to result from incomplete lineage sorting. It has earlier been demonstrated that the pDNA haplotype composition of *B*. *pubescens* changes according to the species occurring in sympatry, which suggests hybridisation rather than incomplete linage sorting as the main reason for shared pDNA (Palmé *et al.* 2004). Similar extensive hybridisation and introgression among sympatric species have been demonstrated in broad analyses of several European oaks (*Quercus*; Dumolin-Lapegue *et al.* 1997; Petit *et al.* 2002).

In contrast, the Bayesian and NJ clustering of our *Betula* ssp. dataset based on mainly nuclear AFLP data suggested genetic groups corresponding to species and subspecies (Fig 2; Fig. S3, Supporting Information). Thus, the AFLP markers, as also found for nuclear microsatellites (Wang *et al.* 2014), seem to be little influenced by introgression compared with plastid markers. Although we identified some admixture, the clear differentiation between *B. pubescens* and *B. nana* at AFLP loci, but not in plastid DNA, implies that hybridisation between them typically has led to plastid capture in the absence of extensive nuclear introgression. Simulations have also

shown that phylogenetic reconstructions based on markers with high levels of introgression, such as pDNA, are likely to be incongruent with the history of the majority of genes that a given organism possesses (Currat *et al.* 2008). Thus, we suggest that phylogeographic inference in *Betula* should be drawn on the basis of nuclear variation rather than pDNA variation when discrepancies are identified.

Although our AFLP data for *B. nana* and *B. pubescens* seem to be less influenced by introgression between the two than are pDNA markers, there is still a possibility that other *Betula* species found sympatrically or at their range margins may have had an influence. We were not able to include all relevant *Betula* species in our analyses, but based on the reference samples included and the range-wide sampling in this study we are confident that our AFLP data mainly reflect species-specific patterns. In *B. nana,* we initially expected to detect some influence from its close relative *B. glandulosa*, but these species were strongly differentiated at AFLP loci, and the only pDNA haplotype shared was the assumed ancient, widespread pDNA haplotype Aa (Fig. 6; Table 1). *Betula glandulosa* also showed distinctly higher levels of genetic rarity than *B. nana* and *B. pubescens*, suggesting that potential hybridisation with *B. glandulosa* has not made any major influence on the genetic patterns seen in the other two species. It is also worth noting that the AFLP data do not support the latest classification by Ashburner & McAllister (2013), where *B. nana ssp. exilis* is referred to as *B. glandulosa*.

The history of *B. nana*

In *B. nana*, the geographic distribution of different pDNA halotypes was associated with different AFLP groups, and largely congruent with the limited pDNA data previously reported for this species (Fig. 4 c; Palmé *et al.* 2004; Maliouchenko *et al.* 2007; Thórsson *et al.* 2010). This

congruence, in combination with the phylogenetic pattern (Fig S3), suggests that interspecific hybridisation has had little influence on genetic structure in *B. nana*, and that refugial isolation and postglacial migration history are the major drivers behind the observed genetic patterns in this species.

The analyses of the *Betula* spp. AFLP dataset identified two genetic groups (hereafter referred to as main groups) in *B. nana*, "*B. nana* East" and "*B. nana* West", largely corresponding to the two partly allopatric subspecies (Fig. 2a). The analyses of the *B. nana* dataset that contained more markers refined this pattern, identifying five genetic groups (hereafter referred to as subgroups) along the east-west gradient. This pattern suggests a hierarchical genetic structure within *B. nana*. The main east-west genetic pattern that also is recognized taxonomically seems to represent an older differentiation process, while the substructuring probably represents more recent processes (late-glacial to postglacial; discussed below). The disjunct distribution of the assumed ancient haplotype Aa, with centres in the east (Beringia – East Siberia) and west (Europe), also supports a hierarchical genetic structure (Fig. 4 c).

Pliocene macrofossils from e.g. North Greenland (Bennike & Böcher 1990) and pollen records from Russia (Andreev *et al.* 2014) suggest that *B. nana* had a widespread, continuous distribution before the onset of the Pleistocene glaciations. The recurrent ice-sheet formation in Europe and West Siberia and the extremely dry glacial conditions in East Siberia (Astakhov 2008; Binney *et al.* 2009) must repeatedly have fragmented the Eurasian distribution of *B. nana*. The areas furthest east, i.e. Beringia, remained unglaciated throughout the Pleistocene, and were not as dry as East Siberia (Elias & Brigham-Grette 2013). Persistence over several glaciations in Beringia may explain the eastern *B. nana* main group, and the Beringia-East Siberia subgroup may

represent both persistent populations and postglacial expansion from this refugium. Beringia is a well-documented refugium for many arctic-alpine plants (Abbott & Brochmann 2003), and pollen records show presence of *B. nana* prior to and during the LGM (Goetcheus & Birks 2001; Ager 2003). Persistence is also in line with the indicated, although not significant, elevated genetic rarity in the Beringian region (Fig. S7a, Supporting Information).

The western main group dominates the Atlantic-European part of the distribution range, which was repeatedly glaciated. Individuals assigned to this group are probably descendants from western population(s), which recurrently expanded and contracted throughout the Pleistocene. Many late-glacial macrofossils have been recorded in the areas south of the Scandinavian ice sheet (Tralau 1963; Hultén & Fries 1986; Binney *et al.* 2009), and *B. nana* is still present in central and southern European mountains today. Thus, *B. nana* may have survived glacial periods in a large, more or less connected glacial refugium along the southern to southeastern glacial margin, and the European subgroup probably represents postglacial descendants from this refugium. However, contrary to expectations (Ibrahim *et al.* 1996; Hewitt 2000; Excoffier *et al.* 2009), we found that genetic diversity increased northwards in Europe. Under the current climate conditions, the cold-adapted *B. nana* has most of its suitable habitat in the north, and it has probably lost suitable habitat and genetic diversity at the rear-edge of its current range (Hampe & Petit 2005; Wang *et al.* 2014). In addition, several populations in Northeast Europe were mixed with individuals belonging to other subgroups, like the West Siberian subgroup, indicating some mixture of expanding postglacial linages, which also may explain higher genetic diversity in the north.

Individuals belonging to the West Siberian subgroup seem to be descendants from initial admixed populations between the main eastern and western groups (Fig. 2 a, b). Refugial isolation during the last glaciation can explain the current differentiation of these admixed populations. The presence of a pDNA haplotype only found in this region (Ca), supports refugial isolation in Western Siberia (Fig. 4 a, c). Evidence for refugia in West Siberia during the last glaciation has been found for several other arctic plant species e.g. (Skrede *et al.* 2006; Eidesen *et al.* 2007a), and is in line with the findings of previous pDNA studies in *Betula* (Palmé 2003; Palmé *et al.* 2004; Maliouchenko *et al.* 2007).

The Svalbard and the Atlantic subgroups contained less genetic diversity, and most likely result from founder effects and bottlenecking during postglacial colonisation rather than refugial isolation in these heavily glaciated areas. Svalbard was heavily glaciated during LGM (Landvik *et al.* 1998), and the only known ice-free area could not sustain growth of *B. nana* (Landvik *et al.* 2003; Alsos *et al.* 2007). Svalbard was probably colonised during the Holocene warm period, and later population size reduction in response to the late Holocene cooling (Birks 1991), would reduce the intrapopulation diversity further. Our results are in line with earlier analyses suggesting that *B. nana* colonised Svalbard from Russia (Alsos *et al.* 2007). Individuals assigned to the Svalbard group were genetically most similar to those of the West Siberian subgroup (PCO, Fig. S6, Supporting Information), and the pDNA haplotype found in Svalbard dominated East Europe and West Siberia.

The Atlantic subgroup is present in Europe, common in Iceland, and dominant in Greenland. During an early phase of deglaciation, the ancestors of this group probably expanded northward from periglacial populations in the westernmost part of the European refugium and up

along the Norwegian coast, followed by dispersal across the North Atlantic to Iceland and Greenland. Such leading-edge expansion can lead to differentiation through successive funder effects and bottlenecking along the way (Hewitt 2000). The significant reduction in intrapopulation genetic diversity with increasing latitude supports such recent northward expansion, possibly in combination with range-contractions in response to the late Holocene cooling. Further, the derived and species specific pDNA haplotype Ab that was exclusive to this subgroup, also supports such a scenario. Periglacial areas were not suitable for *B. pubescens* during LGM. This haplotype may have moved north and westwards at the expanding front of *B. nana*, and diminished at the rear end. The current distribution of the Atlantic subgroup inferred for *B. nana* barely overlaps with the current distribution of *B. pubescens*, reducing the possibilities for hybridisation and plastid capture into *B. pubescens*. Westward colonisation from NW Europe to Iceland and Greenland is in agreement with the succession of dated fossil records (Fredskild 1991), and has been inferred for numerous species based on genetic and floristic evidence (Alsos et al. in press). Recolonisation of Iceland from a refugium in Western Europe is also concordant with previously published pDNA data (Palmé *et al.* 2004; Thórsson *et al.* 2010).

The history of *B. pubescens*

In contrast to *B. nana*, we found no geographical structuring of the AFLP variation in *B. pubescens*, weak isolation by distance, and only weak differentiation among populations. Population diversity was high and showed no reduction in previously glaciated areas. Thus, our mainly nuclear AFLP data suggest broad-fronted, efficient recolonization of *B. pubescens* after the last glaciation from a single main European refugium, or possibly consisting of several smaller but partly connected refugia located rather close to the ice-margin. This is in line with paleobotanical evidence, where macrofossils and pollen influx values for *B.* cf. *pubescens* illustrate rapid postglacial spread

(MacDonald *et al.* 2000) from Central European refugia (Willis *et al.* 2000; Birks & Willis 2008). *Betula* cf. *pubescens* was, for example, the first tree that colonized the Atlantic coast of Norway after the last glaciation, and about 600 years were needed for *Betula* to spread round the Norwegian coast and up the valleys into the mountains (Birks *et al.* 2005). Swedish megafossil evidence has suggested scattered refugia close to the LGM ice margin (Kullman 2002, 2008), although these results have been questioned (Birks *et al.* 2005). The total absence of large-scale structure based on AFLP data and overall high diversity in *B. pubescens* also suggest high levels of current gene flow, both through seeds and pollen, which is in line with local studies from e.g. Sweden and Iceland (Truong *et al.* 2007; Thórsson *et al.* 2010).

The shared pDNA haplotypes (Aa, Ca, and Cc) showed a similar geographical distribution as in *B. nana*, and the overall pDNA pattern we observed in *B. pubescens* was in line with an east-west pattern reported earlier (Fig 4; Palmé 2003; Palmé *et al.* 2004; Maliouchenko *et al.* 2007). The congruent pDNA structure between *B. nana* and *B. pubescens*, but clear discrepancy between pDNA and AFLP data for *B. pubescens*, suggest asymmetric hybridisation and plastid capture from *B. nana* into *B. pubescens* during postglacial range expansion. Macrofossils and pollen data show that *B. pubescens* was rapidly expanding into areas already occupied by *B. nana* (van Dinter & Birks 1996; Bergman *et al.* 2005), and analysis of size distribution of *Betula* pollen diameters from Iceland show increased hybrid formation with *B. nana* during periods of expansion of *B. pubescens* (Karlsdóttir *et al.* 2014). As introgression mainly occurs from the local to the invading species (Currat *et al.* 2008), and more readily from diploids to tetraploids than *vice versa* (Stebbins, 1971; Chapman & Abbott, 2010), introgression during postglacial expansion is expected from the early coloniser and diploid *B. nana* into the later coloniser and tetraploid *B. pubescens*. Crossing experiments in *Betula* also suggest that hybridisation is more successful when the female parent is of low ploidy (Eriksson & Jonsson 1986; DeGroot *et al.* 1997). The tetraploid *B.*

pubescens should therefore be more successful as pollen donor, and the diploid *B. nana* should be more successful as maternal parent. As hybrids tend to backcross with *B. pubescens* (Elkington 1968; Wang *et al.* 2014), *B. pubescens* will be inclined to accumulate plastids in interspecific crosses between species. Our admixture data also supports asymmetric backcrossing of hybrids with *B. pubescens*, particularly at the leading edge of expansion.

The observed leading-edge hybridisation with *B. nana* may also have facilitated the rapid recolonization by *B. pubescens*. Hybrids may have better invasion ability because of increased genetic variation and evolutionary novelty (Ellstrand & Schierenbeck 2006). *Betula pubescens* is less cold-tolerant than *B. nana*, but hybrids between them are likely to be more cold-tolerant than *B. pubescens*. In addition, colonization of new territories may be facilitated by "pollen swamping", allowing some species to settle in an area through initial extensive pollen flow and hybridisation without initial colonization by seeds (Petit *et al.* 2004). When seeds manage to establish later on, there will already by an extensive presence of the invading species' genome in the population through the already established hybrids. These hybrids represent potential mates for the later colonising plants. As a tree species, *B. pubescens* produces more pollen, and is a more efficient pollen disperser than the dwarf-shrub *B. nana*. Thus, *B. pubescens* is probably a more successful pollen donor (Eriksson & Jonsson 1986) and the establishment of hybrids at the leading-edge of expansion through pollen swamping seems likely.

Conclusions

In contrast to the extensive sharing of pDNA haplotypes between *B. nana* and *B. pubescens* shown in this and previous studies (Palmé *et al.* 2004; Maliouchenko *et al.* 2007; Thórsson *et al.* 2010), the two species were clearly differentiated at AFLP loci, allowing phylogeographical inferences to be obtained. *Betula nana* showed a clear east-west structure throughout Eurasia, and the

distribution of pDNA haplotypes was associated with the structuring of the AFLP variation. The AFLP structure in *B. nana* indicates at least two (Europe and Beringia) and possibly three separate refugial areas existed for this species during the last glacial period. These were south of the North European ice sheets, east of the North European ice sheets, and in Beringia. In contrast, *B. pubescens* showed a pDNA pattern similar to that of *B. nana*, but no structuring based on AFLPs. The lack of AFLP structure suggests efficient expansion from a single large refugium combined with historical and current gene flow. The incongruence between pDNA and AFLP variation in *B. pubescens* can be explained by asymmetrical hybridisation with *B. nana*, where hybrids mainly are established with *B. pubescens* as pollen donor, and backcrossed with *B. pubescens*, leading to plastid capture from *B. nana* to *B. pubescens*. We identified a weak but significant increase in nuclear (AFLP) gene flow from *B. nana* into *B. pubescens* with increasing latitude, suggesting hybridisation has been most frequent at the postglacial expansion front of *B. pubescens.* We suggest that extensive pollen flow (pollen swamping) by *B. pubescens* into territories already occupied by the more cold-tolerant *B. nana*, created a zone of more cold-tolerant hybrids, which facilitated reproduction and further expansion for the first *B. pubescens* individuals that later managed to establish through seeds.

Acknowledgements

We thank all collectors (cf. Table S1, Supporting Information), Virginia Mirré and Gro H. Jacobsen for assistance in the lab, Inger Skrede, Dorothee Ehrich and Rémy J. Petit for commenting on an early version of the manuscript, and Tor Carlsen for helpful discussions. STRUCTURE analyses were run at the Lifeportal, University of Oslo. The study was supported by grants 150322/720 and 146515/420 to Christian Brochmann from the Research Council of Norway.

References

Abbott RJ, Brochmann C (2003) History and evolution of the arctic flora: in the footsteps of Eric Hultén. *Molecular Ecology* **12**, 299-313.

Ager TA (2003) Late Quaternary vegetation and climate history of the central Bering land bridge from St. Michael Island, western Alaska. *Quaternary Research* **60**, 19-32.

Alsos IG (2003) *Conservation biology of the most thermophilous plant species in the Arctic: Genetic variation, recruitment and phylogeography in a changing climate* PhD thesis, University of Tromsø.

Alsos IG, Ehrich D, Thuiller W*, et al.* (2012) Genetic consequences of climate change for northern plants. *Proceedings of the Royal Society B-Biological Sciences* **279**, 2042-2051.

Alsos IG, Eidesen PB, Ehrich D*, et al.* (2007) Frequent long-distance plant colonization in the changing Arctic. *Science* **316**, 1606-1609.

Anamthawat-Jónsson K (2012) Hybridisation, introgression and phylogeography of Icelandic Birch. In: *Current topics in phylogenetics and phylogeography of terrestrial and aquatic systems* (ed. Anamthawat-Jónsson K), pp. 117-144. InTech, www.intechopen.com.

Andreev AA, Tarasov PE, Wennrich V*, et al.* (2014) Late Pliocene and Early Pleistocene vegetation history of northeastern Russian Arctic inferred from the Lake El'gygytgyn pollen record. *Climate of the Past* **10**, 1017-1039.

Ashburner K, McAllister HA (2013) *The genus Betula. A taxonomic revision of birches* Royal Botanic Gardens, Kew.

Astakhov V (2008) Geographical extremes in the glacial history of northern Eurasia: post-QUEEN considerations. *Polar Research* **27**, 280-288.

Atkinson MD (1992) *Betula pendula* Roth (B-Verrucosa Ehrh) and *B pubescens* Ehrh. *Journal of Ecology* **80**, 837-870.

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37-48.

Bennike O, Böcher J (1990) Forest-tundra neighboring the north pole: plant and insect remains from the Plio-Pleistocene Kap København formation, North Greenland. *Arctic* **43**, 331-338.

Bergman J, Hammarlund D, Hannon G, Barnekow L, Wohlfarth B (2005) Deglacial vegetation succession and Holocene tree-limit dynamics in the Scandes Mountains,

west-central Sweden: stratigraphic data compared to megafossil evidence. *Review of Palaeobotany And Palynology* **134**, 129-151.

Binney HA, Willis KJ, Edwards ME*, et al.* (2009) The distribution of late-Quaternary woody taxa in northern Eurasia: evidence from a new macrofossil database. *Quaternary Science Reviews* **28**, 2445-2464.

Birks HH (1991) Holocene vegetational history and climatic change in west Spitsbergen - plant macrofossils from Skardtjorna, an Arctic lake. *The Holocene* **1**, 209-218.

Birks HH, Larsen E, Birks HJB (2005) Did tree-*Betula*, *Pinus* and *Picea* survive the last glaciation along the west coast of Norway? A review of the evidence, in light of Kullman (2002). *Journal of Biogeography* **32**, 1461-1471.

Birks HJB, Willis KJ (2008) Alpines, trees, and refugia in Europe. *Plant Ecology & Diversity* **1**, 147-160.

Bonin A, Bellemain E, Eidesen PB*, et al.* (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* **13**, 3261-3273.

Chapman MA, Abbott RJ (2010) Introgression of fitness genes across a ploidy barrier. *The New phytologist* **186**, 63-71.

Corander J, Waldmann P, Sillanpaa MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics* **163**, 367-374.

Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. *Evolution* **62**, 1908-1920.

DeGroot WJ, Thomas PA, Wein RW (1997) *Betula nana* L. and *Betula glandulosa* Michx. *Journal of Ecology* **85**, 241-264.

Dumolin-Lapegue S, Demesure B, Fineschi S, Corre VL, Petit RJ (1997) Phylogeographic structure of white oaks throughout the European continent. *Genetics* **146**, 1475-1487.

Dyke AS (2004) An outline of North American deglaciation with emphasis on central and northern Canada. In: *Quaternary Glaciations: Extent and Chronology* (eds. Ehlers J, Gibbard PL), pp. 373-424. Elsevier, Amsterdam.

Earl, D, vonHoldt, BM (2012) Structure harvester: a website and program for visualizing Structure output and implementing the Evanno method. Conservation Genetics Resources **4**, 359-361

Ehrich D (2006) aflpdat: a collection of r functions for convenient handling of AFLP data. *Molecular Ecology Notes* **6**, 603-604.

Eidesen PB, Alsos IG, Popp M*, et al.* (2007a) Nuclear vs. plastid data: complex

Pleistocene history of a circumpolar key species. *Molecular Ecology* **16**, 3902-3925.

Eidesen PB, Carlsen T, Molau U, Brochmann C (2007b) Repeatedly out of Beringia: *Cassiope tetragona* embraces the Arctic. *Journal of Biogeography* **34**, 1559-1574.

Eidesen PB, Ehrich D, Bakkestuen V*, et al.* (2013) Genetic roadmap of the Arctic: plant dispersal highways, traffic barriers and capitals of diversity. *New Phytologist* **200**, 898–910.

Elias SA, Brigham-Grette J (2013) Glaciations: Late Pleistocene Glacial Events in Beringia. In: *Encyclopedia of Quaternary Science* (ed. Elias S), pp. 191-201. Elsevier, Amsterdam.

Elkington TT (1968) Introgressive hybridization between *Betula nana* L and *B Pubescens* Ehrh In North-West Iceland. *New Phytologist* **67**, 109-118.

Ellstrand NC, Schierenbeck KA (2006) Hybridization as a stimulus for the evolution of invasiveness in plants? *Euphytica* **148**, 35-46.

Elven R, Murray DF, Razzhivin V, Yurtsev BA (2011) *Checklist of the Panarctic Flora (PAF): Vascular plants, Draft version.* Natural History Museum, University of Oslo, http://nhm2.uio.no/paf/

Eriksson G, Jonsson A (1986) A review of the genetics of *Betula*. *Scandinavian Journal of Forest Research* **1**, 421-434.

Ersts PJ (Internet) *Geographic Distance Matrix Generator(version 1.2.3).* American Museum of Natural History, Center for Biodiversity and Conservation, http://biodiversityinformatics.amnh.org/open_source/gdmg

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

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resourses* **10**, 564-567.

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

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

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

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* **7**, 574-578.

Fredskild B (1991) The genus *Betula* in Greenland-Holocene history, present distribution and synecology. *Nordic Journal of Botany* **11**, 393-412.

Furlow JJ (1997) *Betula*. In: *Flora of North America North of Mexico.* Vol. 3, New York and Oxford.

Goetcheus VG, Birks HH (2001) Full-glacial upland tundra vegetation preserved under tephra in the Beringia National Park, Seward Peninsula, Alaska. *Quaternary Science Reviews* **20**, 135-147.

Hamilton MB (1999) Primer Notes - Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology* **8**, 521-523.

Hammer Ø, Harper DAT, Ryan PD (2001) PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica* **4**, 9.

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

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907-913.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* **9**, 1322-1332.

Hultén E, Fries M (1986) *Atlas of North European Vascular Plants North of the Tropic of Cancer* Koeltz Scientific Books, Königstein.

Ibrahim KM, Nichols RA, Hewitt GM (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity* **77**, 282-291.

Jadwiszczak KA (2012) What can molecular markers tell us about the glacial and postglacial histories of European birches? *Silva Fennica* **46**, 733–745.

Jadwiszczak KA, Banaszek A, Chrzanowska A, Kłosowski S, Sozinov OV (2014) The admixture zone of Betula humilis Schrk. phylogenetic lineages follows the eastern central European suture zone. *Plant Ecology & Diversity*, 1-7.

Jadwiszczak KA, Banaszek A, Jabłońska E, Sozinov OV (2012) Chloroplast DNA variation of *Betula humilis* Schrk. in Poland and Belarus. *Tree Genetics & Genomes* **8**, 1017-1030.

Jonsell B, Karlsson T (2000) *Flora nordica* Bergius Foundation, Stockholm.

Järvinen P, Palmé A, Orlando Morales L*, et al.* (2004) Phylogenetic relationships of *Betula* species (Betulaceae) based on nuclear ADH and chloroplast matK sequences. *American Journal of Botany* **91**, 1834-1845.

Karlsdóttir L, Hallsdóttir M, Eggertsson O, Thorsson AET, Anamthawat-Jonsson K (2014) Birch hybridization in Thistilfjördur, North-east Iceland during the Holocene. *Icelandic Agricultural Sciences* **27**, 95-109.

Kosman E (2003) Nei's gene diversity and the index of average differences are identical measures of diversity within populations. *Plant Pathology* **52**, 533-535.

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

Kullman L (2008) Early postglacial appearance of tree species in northern Scandinavia: review and perspective. *Quaternary Science Reviews* **27**, 2467-2472.

Landvik JY, Bondevik S, Elverhøi A*, et al.* (1998) The last glacial maximum of Svalbard and the Barents Sea area: Ice sheet extent and configuration. *Quaternary Science Reviews* **17**, 43-75.

Landvik JY, Brook EJ, Gualtieri L*, et al.* (2003) Northwest Svalbard during the last glaciation: Ice-free areas existed. *Geology* **31**, 905-908.

MacDonald GM, Velichko AA, Kremenetski CV*, et al.* (2000) Holocene treeline history and climate change across Northern Eurasia. *Quaternary Research* **53**, 302-311.

Maliouchenko O, Palmé AE, Buonamici A, Vendramin GG, Lascoux M (2007) Comparative phylogeography and population structure of European *Betula* species, with particular focus on *B. pendula* and *B. pubescens*. *Journal of Biogeography* **34**, 1601-1610.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, 209-220.

Nei M (1987) *Molecular Evolutionary Genetics* Columbia University Press, New York.

Palmé AE (2003) *Evolutionary history and chloroplast DNA variation in three plant genera: Betula, Corylus and Salix. The impact of post-glacial colonization and hybridization* PhD thesis, Uppsala University.

Palmé AE, Su Q, Palsson S, Lascoux M (2004) Extensive sharing of chloroplast haplotypes among European birches indicates hybridization among *Betula pendula*, *B. pubescens* and *B. nana*. *Molecular Ecology* **13**, 167-178.

Petit RJ, Bodenes C, Ducousso A, Roussel G, Kremer A (2004) Hybridization as a mechanism of invasion in oaks. *New Phytologist* **161**, 151-164.

Petit RJ, Csaikl UM, Bordacs S*, et al.* (2002) Chloroplast DNA variation in European white oaks: Phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management* **156**, 5-26.

Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution* **16**, 37-45.

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

R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. , http://www.R-project.org/.

Rieseberg LH, Soltis DE (1991) Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends In Plants* **5**, 65-84.

Rieseberg LH, Whitton J, Linder CR (1996) Molecular marker incongruence in plant hybrid zones and phylogenetic trees. *Acta Botanica Neerlandica* **45**, 243-262.

Schenk MF, Thienpont C-N, Koopman WJM, Gilissen LJWJ, Smulders MJM (2008) Phylogenetic relationships in *Betula* (Betulaceae) based on AFLP markers. *Tree Genetics & Genomes* **4**, 911-924.

Schönswetter P, Tribsch A (2005) Vicariance and dispersal in the alpine perennial *Bupleurum stellatum* L. (Apiaceae). *Taxon* **54**, 725-732.

Shaw J, Lickey EB, Beck JT*, et al.* (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* **92**, 142-166.

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany* **94**, 275-288.

Shaw KL (2002) Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 16122-16127.

Skrede I, Eidesen PB, Portela RP, Brochmann C (2006) Refugia, differentiation and postglacial migration in arctic-alpine Eurasia, exemplified by the mountain avens (*Dryas octopetala* L.). *Molecular Ecology* **15**, 1827-1840.

Stebbins GL (1971) *Chromosomal evolution in higher plants*. Edward Arnold, London, UK

Svendsen J, Alexanderson H, Astakhovc VI, *et al.* (2004) Late Quaternary ice sheet history

of northern Eurasia. *Quaternary Science Reviews* **23**, 1229-1271.

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17**, 1105-1109.

Thomson AM, Dick CW, Dayanandan S, Carine M (2015) A similar phylogeographical structure among sympatric North American birches (*Betula*) is better explained by introgression than by shared biogeographical history. *Journal of Biogeography* **42**, 339–350.

Thórsson ÆT, Pálsson S, Lascoux M, Anamthawat-Jónsson K (2010) Introgression and phylogeography of *Betula nana* (diploid), *B. pubescens* (tetraploid) and their triploid hybrids in Iceland inferred from cpDNA haplotype variation. *Journal of Biogeography* **37**, 2098-2110.

Thórsson ÆT, Salmela E, Anamthawat-Jónsson K (2001) Morphological, cytogenetic, and molecular evidence for introgressive hybridization in birch. *Journal of Heredity* **92**, 404-408.

Tralau H (1963) Distribution of boreal and arctic montane plants. *Arkiv för Botanik* **5**, 533-582.

Truong C, Palme AE, Felber F (2007) Recent invasion of the mountain birch *Betula pubescens* ssp. *tortuosa* above the treeline due to climate change: genetic and ecological study in northern Sweden. *Journal of Evolutionary Biology* **20**, 369-380.

Tutin TG (1964) *Flora Europaea* Cambridge University Press, Cambridge.

van Dinter M, Birks H (1996) Distinguishing fossil *Betula nan*a and *B. pubescens* using their wingless fruits: implications for the late-glacial vegetational history of western Norway. *Vegetation History and Archaeobotany* **5**, 229-240.

Vos P, Hogers R, Bleeker M*, et al.* (1995) AFLP - a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**, 4407-4414.

Wang N, Borrell JS, Bodles WJ*, et al.* (2014) Molecular footprints of the Holocene retreat of dwarf birch in Britain. *Molecular Ecology* **23**, 2771-2782.

Willis KJ, Rudner E, Sumegi P (2000) The full-glacial forests of Central and Southeastern Europe. *Quaternary Research* **53**, 203-213.

Data Accessibility

The original AFLP matrices, the three AFLP data matrixes used for the data presented here, the geographical distance matrices used for mantel tests, the marker-frequency matrices used in NJ

analyses among populations, input matrices for STRUCTURE analyses, and alignments used to generate haplotypes (Table S3) are available in Dryad doi:10.5061/dryad.c3g80, and sequence data have been deposited in GenBank under accession nos. KR779549 - KR779606 and KR779607 - KR779658.

Author Contributions

P. B. Eidesen developed the idea, contributed to sampling, led the lab work, analysed the data and led the writing; I.G. Alsos contributed to back ground information and sampling, discussed the results, and commented on draft; C. Brochmann conceived the ideas, provided funding, discussed the results and participated in writing.

Tables

Table 1. Plastid DNA haplotypes observed in species of *Betula* based on combined sequences (18 variable sites) of *trn*S-*trn*G (variants denoted by uppercase letters) and *trn*Q-*rps*16 (variants denoted by lowercase letters; see Table S3, Supporting Information for details). *Betula michauxii* and *B. pendula* were not successfully analysed for *trn*Q-*rps*16, but their *trn*S-*trn*G variant are indicated. Two hybrids between *B. nana* and *B. pubescens* were included, containing haplotype Ca and Cc, and these were counted as 0.5 to each of the parental species.

Haplotype	Aa	Ab	Af	A-	Be	Ca	Cc	Dd	Ed	Fe	Ge	Gg	Gh	G-	Ha	Ia
B. nana	9	7				1.5	6.5	1								2
B. pubescens	5				3	1.5	3.5		1	1					1	
B. glandulosa	1		1													
B. pendula				1												
B. pumila											1					
B. michauxii														1		
B. cf. *minor*												1	1			
Total	15	7	1	1	3	3	10	1	1	1	1	1	1	1	1	2

Figure legends

Fig. 1 a) Geographical distribution of *Betula nana*, *B. glandulosa* and *B. pubescens* based on Hultén & Fries (1986), Furlow (1997) and Elven *et al.* (2011).. b) Approximate extent of glaciers during the last glacial maximum (LGM, about 20 000 yr ago) based on Dyke (2004) and Svendsen *et al.* (2004).

Fig. 2 Genetic structure based on STRUCTURE analyses of multilocus AFLP phenotypes. a) Genetic structure based on the complete dataset with 901 samples of *Betula* scored for 104 AFLP markers (best fit for $K = 4$). b) Genetic structure in *B. nana* (528 samples based on three primer combinations and 115 markers that were polymorphic in this species; best fit for $K = 5$). Increasing the number of polymorphic markers in *B. pubescens* did not reveal any further substructure (not shown).

Fig. 3 Trend of admixture in relation to latitude a) from *B. nana* into *B. pubescens* and b) from *B. pubescens* into *B. nana*, inferred from logit-transformed STRUCTURE admixture proportions of multilocus AFLP phenotypes. Black circles indicate values per individual. Red diamonds represent values for each population fitted by the mixed effects model given by Wang *et al.* (2014), with slope as a fixed effect and population modelled as a random effect to allow for genetic drift. In *B. pubescens* (a), levels of admixture significantly increase with latitude (n=303; slope 0.063, SE 0.016, $p < 0.001$), while no such trend was found in *B. nana* (n=531; slope 0.000, SE 0.018, $p = 0.999$).

Fig. 4 Geographical distribution of AFLP (a, b) and plastid DNA (c, d) variation in *Betula nana* and *B. pubescens*. Colours in a) and b) (cf. Fig. 2) identify genetic groups inferred from STRUCTURE analyses of multilocus AFLP phenotypes. Coloured symbols and pie charts in c) and d) show distribution of pDNA variation. Coloured symbols identify combined plastid haplotypes based on sequencing of the *trn*S-*trn*G and *trn*Q-*rps*16 regions in this study (Table S3, Supporting Information), with colour coding according to haplotype groups (cf. Fig. 6). Pie charts show the distribution of plastid haplotype groups identified in previous studies, redrawn after Maliouchenko *et al.* (2007). Resemblance of colours among haplotype groups presented in pie charts (Maliouchenko *et al.*, 2007) and our data indicate similar geographical distribution of haplotype groups in both studies, but the actual similarity among haplotypes are unknown, as they are retrieved from different pDNA marker systems.

Fig. 5 Geographical distribution of intrapopulation genetic diversity in a) *Betula nana* and b) *B. pubescens* based on multilocus AFLP phenotypes generated with three primer combinations (D3;

Table S2, Supporting Information).

Fig. 6 Plastid DNA haplotype network based on 47 samples mainly representing *Betula nana* (BN) and *B. pubescens* (BP), but also North American reference individuals of *B. glandulosa* (BG), *B. pumila* and *B*. cf. *minor*. Haplotype Gh found in *B*. cf. *minor* was omitted from the network due to missing data. Number of variable sites separating haplotypes is indicated along branches. Haplotypes are based on combined sequences (18 variable sites) of two plastid regions, *trn*S-*trn*G and *trn*Q-*rps*16 (Table S3, Supporting Information). Circle size is proportional to number of individuals.