Ecological Genomics for the Conservation of Dwarf Birch

James S. Borrell

School of Biological and Chemical Sciences,

Queen Mary University of London,

Mile End Road,

London E1 4NS

Supervisors: Dr. Richard J. A. Buggs &
Professor Richard A. Nichols

A thesis submitted to the

University of London

for the degree of

Doctor of Philosophy

November 2016

Declaration

I, James Borrell, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I acknowledge particular data acquisition and analytical assistance as follows: Richard Buggs performed collections of *Betula pubescens* material included in this study, and Nian Wang conducted lab based preparation and sequencing of these samples. Jasmin Zohren produced an assembly of the *Betula nana* genome, which was a valuable tool throughout this research. Jasmin Zohren also produced the RNAseq data discussed in Chapter 4, and assisted in formatting her data for analysis of bidirectional introgression. Richard Nichols helped write scripts to perform maximum likelihood F_{ST} analysis discussed in Chapter 3.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis. I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university. The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Publications

The research presented in this thesis has directly contributed to the following publications:

Borrell, J.S., Wang, N., Nichols R.A. & Buggs, R.J.A. 2017. Combining markers with different mutation rates for population genetic inference in fragmented tree populations. Evolutionary Applications. *In review*.

Borrell, J.S., Zohren, J., Wang, N., Nichols R.A. & Buggs, R.J.A. 2017. Assisted gene flow can maintain species' adaptive potential. Ecology Letters. *In review*.

Zohren, J., Wang, N., Kardailsky, I., Borrell, J.S., et al., 2016. Unidirectional diploid-tetraploid introgression among British birch trees with shifting ranges shown by RAD markers. *Molecular Ecology*, pp.2413–2426.

Wang, N. et al., 2014. Molecular footprints of the Holocene retreat of dwarf birch in Britain. *Molecular Ecology*, 23(11), pp.2771–2782.

Wang, N., Borrell, J.S., Buggs, R.J.A. 2014b. Is the Atkinson discriminant function a reliable method for distinguishing between Betula pendula and B. pubescens? *New Journal of Botany* **4**: 90-94.

Collaborations and Funding

This research was funded by a NERC CASE studentship NE/J017388/1 in collaboration with conservation NGOs Trees for Life and Highland Birchwoods. Fieldwork in Scandinavia was funded by an EU-INTERACT transnational access grant, which facilitated fieldwork at Kevo subarctic research station, Finland. I gratefully acknowledge the assistance provided by members of the Montane Scrub Action Group and numerous landowners who permitted access to their estates.

Acknowledgements

Firstly, I wish to thank my PhD supervisor, Dr. Richard Buggs, under whose guidance I have been able to explore the aspects of ecological genomics that most interest me. I have benefitted also from numerous interesting discussions with my co-supervisor Prof. Richard Nichols, whom encouraged me to think more deeply about genetics in general.

For fieldwork support I wish to thank Billy Bodles who advised me, accommodated me, and was tireless in his efforts to find these particularly elusive trees. I thank also Alan Featherstone Watson, Jill Hodge, Mick Drury and the staff at Trees for Life, whose work will undoubtedly change the highland landscape forever. I thank Diana Gilbert whose own research on Dwarf birch gave me a head start for this project.

I wish to thank all of the students and faculty in the School of Biological and Chemical Sciences who have made the entire experience enjoyable, and have been quick to share ideas, knowledge and skills. In particular: Hannes Becher, Jasmin Zohren, Nian Wang, Rodrigo Pracana, Steven Dodsworth, Maite Guignard, Bob Verity, Laura Kelly, Adrian Larkeryd, Bruno Viera, Andrea Hatlen and numerous others.

Lastly, I would like to thank my friends, family and partner Jess, who I think are still not quite sure exactly what it is I have been doing for the past four years. I hope this thesis will provide a hefty paperweight for decades to come. I couldn't have completed it without your encouragement.

Abstract

The persistence of woody plant populations faces numerous environmental challenges, including climate change, hybridisation and population fragmentation. Here I explore the genomic signatures and relative importance of these pressures in Dwarf Birch (*Betula nana*), which has declined significantly over the last century across the Scottish Highlands. Firstly, I find that future climate is likely to result in a significant range reduction and that relict populations are likely to display reduced fitness. Secondly, I show that combining multiple mutation rate markers yields more accurate estimates of demographic history and the impact of fragmentation. I develop a novel method to derive high mutation rate markers from short sequencing reads, to facilitate more widespread application. Thirdly, I assess the degree of local adaptation, and explore potential for composite provenancing for the restoration of *B. nana* populations. Surprisingly, the data yields little evidence of adaptive introgression from the related tree *B. pubescens*, suggesting that this may not be an alternative route to climate tolerance. Finally, I review published literature on the population structure and genetic diversity of genus *Betula* in Europe and consider options for the conservation and management of *B. nana*, including assisted gene flow and prioritization of *in situ* genetic diversity.

Contents

Abstract 5

Contents 6

List of figures 8

List of tables 10

Chapter 1: General Introduction 11

Summary 11

1.1 A brief history of woody plants in Europe 12

1.2 Threats to the conservation of woody plants in Europe 18

1.3 Genus *Betula* as a study system 28

1.4 Utilizing ecological genomics to assess threats 35

Chapter 2: Evaluating the distribution and fitness of *B. nana* under current and future climate 40

Summary 40

2.1 Introduction 41

2.2 Methods 44

2.3 Results 51

2.4 Discussion 66

Chapter 3: Combining markers with different mutation rates to infer the population genetic consequences of fragmentation in dwarf birch 69

Summary 69

3.1 Introduction 70

3.2 Materials and Methods 74

3.3 Results 82

3.4 Discussion 96

Chapter 4: Genomic assessment of adaptive potential in dwarf birch under present and future climates 103

Summary 103

4.1 Introduction 104

4.2 Materials and Methods 108

4.3 Results 116

4.4 Discussion 124

Chapter 5: A review of the ecological genomics of *Betula*, and other montane/arctic species in Europe 134

Summary 134

5.1 Comparative population structure and genetic diversity of montane trees, dwarf shrubs and perennial plants in Europe 135

5.2 Comparative population structure and genetic diversity of genus *Betula* in Europe 143

5.3 Region wide population structure and genetic diversity of *Betula nana* 147

5.4 Current and future work by CASE partners for the conservation of *B. nana* in the UK 149

5.5 Conclusions and policy summary for CASE partners 155

Chapter 6: Conclusions 157

6.1 The contribution of this thesis 157

6.2 Novel insights for the conservation of dwarf birch 157

6.3 Methodological advances 159

6.4 New questions and future research 160

Supplementary Information 162

References 175

List of figures

Figure 1.1 The European paleoenvironment during the Last Glacial Maximum 13

Figure 1.2 A schematic representation of the distribution of major vegetation types 15

Figure 1.3 Consequences of contact between species ranges. 24

Figure 1.4 *Betula nana* individuals from across the Scottish highlands.......................................29

Figure 1.5 Theoretical zonation of tree species and habitats across Scottish Mountains............32

Figure 2.1 Historical records of dwarf birch in the UK with normalized sampling effort 51

Figure 2.2 Dwarf birch sampling locations in the UK and Scandinavia 52

Figure 2.3 MaxEnt models of *B. nana* and *B. pubescens* under present and future climate.........55

Figure 2.4 Changes in overall habitat suitability under four future climate scenarios 56

Figure 2.5 Occurance of *B. nana* across slope and aspect values in the UK and Scandinavia 58

Figure 2.6 Boxplot of phenotypic traits for *B. nana*...58

Figure 2.7 Scatterplot of plant height and percentage browsing in the UK 61

Figure 2.8 PCA of phenotypic measurements in *B. nana*..62

Figure 2.9 Boxplot of male and female catkin counts across all populations 63

Figure 2.10 GLM of mean catkin counts against MaxEnt habitat suitability index.......................65

Figure 3.1 Diagram of proposed scenatios for the demographic history of dwarf birch after the Last Glacial Maximum 79

Figure 3.2 Scatter plots of expected heterozygosity, inbreeding and private allele frequency for all four marker datasets 89

Figure 3.3 STRUCTURE plot of dwarf birch populations for all markers sets...............................90

Figure 3.4 Demographic density distributions for all marker sets individually and combined......93

Figure 3.5 Mantel test of pairwise F_{ST} and geographic distance for the UK and Scandinavia........94

Figure 3.6 Scatter plot of log effective population size against maximum liklihood F_{ST} 94

Figure 3.7 Regression of rarefied allelic richness against maximum liklihood F_{ST} 95

Figure 4.1 Schematic of 'Deficit in Adaptive Potential'...114

Figure 4.2 Topology of collinearity between environmental variables used in this study 117

Figure 4.3 Structure plot identifying low levels of introgression 118

Figure 4.4 BLUP estimate of environmentally associatied loci 122

Figure 4.5 Correlation between Shapley values and DAP and barplot of Shapley values 123

Figure 4.6 Hypothetical plot of assisted gene flow for dwarf birch in the UK 124

Figure 4.6 Hypothetical plot of assisted gene flow for dwarf birch in the UK 130

Figure 5.1 Geographic distribution of AFLP diversity in *B. nana* and *B. pubescens* 148

Figure 5.2 Distribution of *B. nana* across Trees for Life study area 150

Figure 5.3 Mean B. nana browsing percentage across Trees for Life study area 151

Figure 5.4 Number of catkins against percentag browsing for *B. nana* 154

List of tables

Table 2.1 Bioclimatic and abiotic variables included in this study 48

Table 2.2 Projected change in global mean surface air temperature..50

Table 2.3 Contribution of environmental variables to MaxEnt model 56

Table 2.4 Changes in suitable habitat area for *B. nana* and *B. pubescens*....................................57

Table 2.5 Sampling locations and phenotypic data summary 59

Table 2.6 Summary of seed production and germination data for B. nana in the UK 63

Table 2.7 Niche distance and HSI values under current and future conditions............................64

Table 3.1 Prior parameters used in DIY-ABC analysis 80

Table 3.2 Summary statistics used in DIY-ABC analysis 80

Table 3.3 Sampling locations, altitude and sample size across the UK and Scandinavia 83

Table 3.4 Summary statistics for four different marker sets 85

Table 3.5 Population genetic summary statistics ordered by decreasing latitude 87

Table 3.6 Posterior demographic parameter estimates for each marker type 91

Table 3.7 Maximum liklihood F_{ST} across all markers with effective population size estimate 92

Table 4.1 *Betula pubescens* sampling locations and number of individuals...............................109

Table 4.2 Environmental variables included in this study and association SNPs........................117

Table 4.3 Environmental variable percentage contributions and associated SNPs....................119

Table 4.4 Significantly overrepresented gene ontology terms...120

Table 4.5 Deficit in adaptive potential across all BLUP analyses 121

Table 5.1 Comparison of population structure and genetic diversity in montane trees, dwarf shrubs and perennials..140

Table 5.2 Summary of phylogeographic studies on Betula sp. in Europe...................................147

Chapter 1: General Introduction

Summary

Woody plants are often keystone species underpinning numerous ecosystems and the services derived from them. In Europe, the diversity and distribution of plants has oscillated as a result of repeated glaciations. In this introduction I first summarise the known distribution of tree species during and since the Last Glacial Maximum and discuss theories on the location of glacial refugia. Furthermore, I highlight the importance of receding edge populations and draw inferences on how this has resulted in currently observed patterns of genetic diversity. Secondly, I discuss processes such as habitat fragmentation, climate change and the movement of species that have resulted in increased pressure on tree populations. I briefly summarise the current threats to woody plants in Europe, particularly montane species, with regard to loss of genetic diversity and future anthropogenic pressures including climate change. Thirdly, I introduce the genus *Betula* as an ideal study system in which to assess the genomic signatures of these pressures. Specifically, the dwarf tree *B. nana* is a montane species threatened by severe reduction in effective population size, population fragmentation, climate change and introgression from congeneric species. It benefits from extensive historical records in the UK, small size enabling experimental tractability and the availability of an assembled genome sequence. Finally I outline the distribution modeling and ecological genomics strategies undertaken in this study, to address i) The current and future distribution of *B. nana*. ii) The potential for microsatellite and SNP markers to assess the impact of fragmentation and genetic drift, and iii) The potential for local adaptation and adaptive introgression to influence the evolutionary potential of *B. nana.* I conclude by evaluating these complementary approaches to consider how they might inform and prioritise the conservation of *B. nana* and other threatened woody plants.

1.1 A brief history of woody plants in Europe

The distribution of genetic and species diversity in the Northern Hemisphere has been strongly influenced by climatic processes during the last Quaternary cycle, encompassing several glacials and interglacials from approximately 2.6 million years ago to present (Hewitt, 1999; McGlone, 1996; Raymo, 1994). Present-day distributions have most recently been shaped by glacial retreat since the Last Glacial Maximum (Birks & Willis, 2008; Huntley & Birks, 1983). Across Europe, ice sheets reached their southernmost extent around 22,000 years before present, with estimates in different localities ranging from 26,500 - 19,000bp. Deglaciation was induced by increased northern summer insolation, which occurred concurrently with abrupt sea level rise (Clark et al., 2009). By around 6000 years ago, the vegetation distribution is thought to be broadly similar to the present (Brewer et al. 2002).

The 'classic' paradigm states that in Europe temperate plant species retreated to three lower–latitude refugia, principally the Iberian, Italian and Balkan peninsulas (see Huntley & Birks, 1983). Refugia refer to areas where plants and animals could survive and reproduce during adverse environmental conditions (Provan & Bennett 2008). This is supported by Palaeontological, palynological and modern biogeographical evidence that indicates many temperate species persisted here where climatic conditions were less severe (Bennett et al. 1991; Schönswetter et al. 2005). Subsequently, as the glaciers retreated numerous species expanded from these refugia to recolonize Northern regions (Douda et al. 2014; Cornille et al. 2013; Magri et al. 2006). The most recent synthesis of the European paleoenvironment during the LGM is given in Figure 1.1 (Tzedakis et al. 2013).

1.1.1 Cryptic refugia during the last glaciation

More recently, phylogeographic studies have suggested the existence of nunataks (Birks 1994) or unknown 'cryptic' refugia (Willis et al. 2000; Stewart & Lister 2001; Magri et al. 2006; de Lafontaine et al. 2014). This adds complexity to the traditional view of expansion from the

three major refugia. In contrast to the major Southern refugia, small cryptic refugia at higher latitudes might be envisaged as pockets of environmentally favorable conditions such as sheltered valleys and south-facing slopes. Two possible interpretations of LGM refugia are displayed in Figure 1.2, reproduced from Birks & Willis (2008).

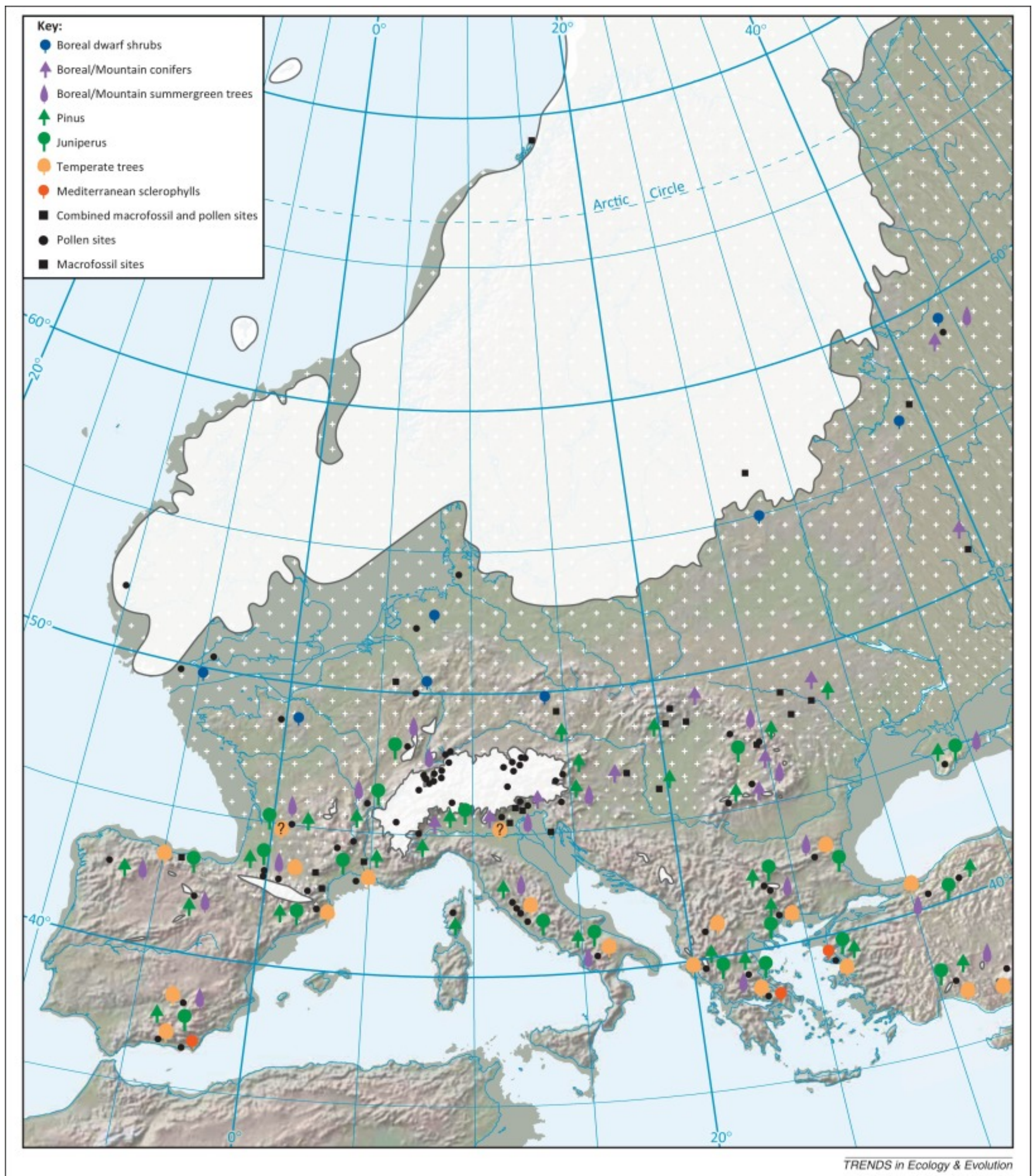


Figure 1.1 The European paleoenvironment during the LGM. Large and small crosses denote continuous and discontinuous permafrost. Boreal dwarf shrubs consist of *Betula nana* and *Salix sp*. Figure reproduced from Tzedakis et al. (2013) under a Creative Commons BY NC ND licence.

Northern refugia have been appealing in order to address Reid's Paradox (Reid 1899; Clark et al. 1998), in which Clement Reid expressed surprise that Oak trees (*Quercus sp*) were already present in Northern Britain, meaning they must have dispersed some 950Km since the end of the LGM far exceeding their predicted dispersal capacity. This paradox may be resolved with the occurrence of rare long-distance dispersal events (Alsos et al. 2007), or the presence of cryptic refugia (Bhagwat & Willis 2008) at higher latitudes that allowed small oak populations to survive, thus reducing the distance they would have to disperse after the ice retreated.

Several factors make reliable identification of cryptic refugia difficult using palynological approaches (de Lafontaine et al. 2014). It is likely that trees could have occurred at population densities too low to create detectable pollen rain in the sediment. Furthermore, some authors suggest that the prevailing climatic conditions would have suppressed pollen production during glacial periods. Finally, it may be difficult to distinguish pollen from small relictual populations to that arriving as a result of long distance dispersal. If higher latitude refugia did indeed exist, then the dispersal capacity of many species may have been overestimated. This has significant implications for our understanding of species' ability to respond to future climate shifts (Kremer et al. 2012; Alberto et al. 2013; Lindner et al. 2008).

Feurdean and authors (2013), for example, found that if Northern refugia are considered then post-glacial migration rates may be revised down substantially. They also found evidence for early successional species such as *Betula*, *Pinus* and *Alnus sp* migrating faster than other species. Migration rates assuming only Southern refugia for these species vary from 225 to 540m yr^{-1} whereas if Northern refugia are considered plausible then rates are revised to 100 to 260m yr-^{1}.

The presence of refugia at higher latitudes remains controversial, despite significant implications for interpreting genetic diversity patterns, migration capacity and conservation strategies (Tzedakis et al. 2013). Reviewing macrofossil, pollen and genetic evidence, Tzedakis and authors found little support for Northern refugia, and thus reject calls to reduce estimates

of migration rates (200-500 m/year^{-1}) by an order of magnitude. Given the difficulty of using palynological data, more recent studies have used explored genetic evidence in an attempt to resolve the issue of Northern refugia and thus accurately interpret current patterns of genetic diversity in Europe.

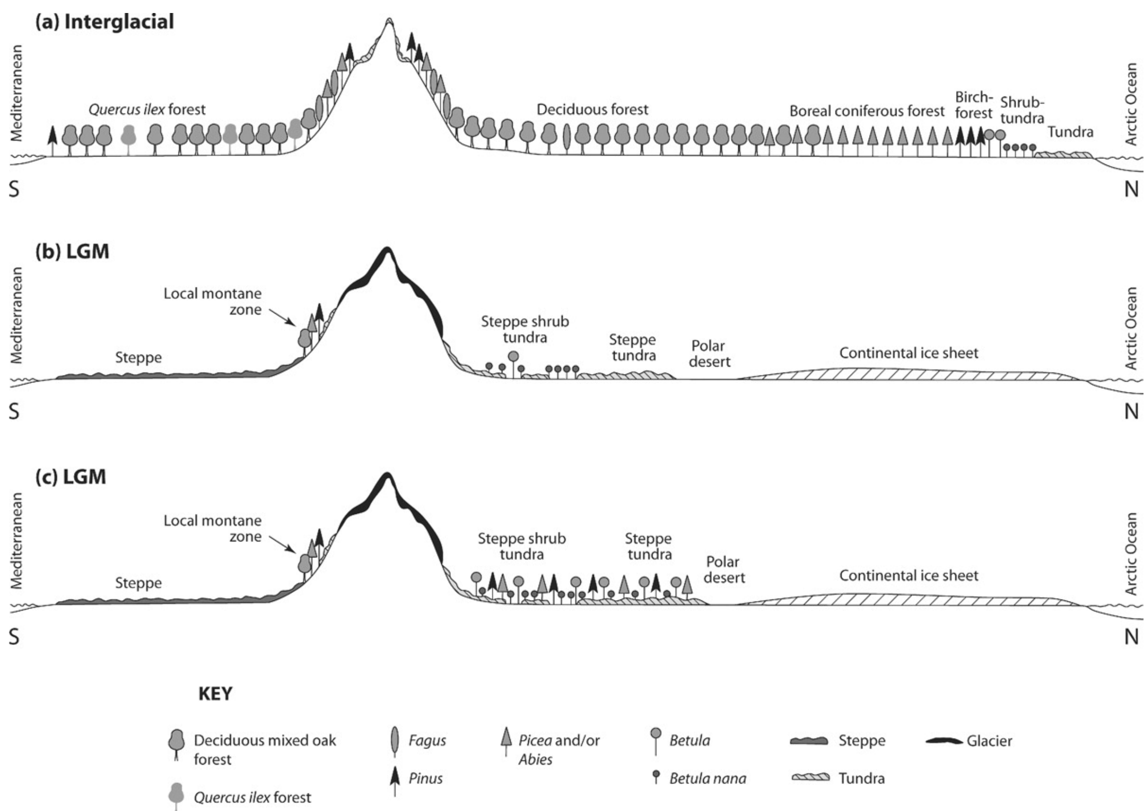


Figure 1.2 A schematic representation of the distribution of major vegetation types. Sequence shows vegetation during a) an interglacial and b) the LGM assuming only southern refugia and c) the LGM assuming both southern and northern refugia. Reproduced with permission from (Birks & Willis 2008).

1.1.2 Genetic patterns of post-glacial expansion

The contraction and expansion of species from glacial refugia should be evidenced in patterns of genetic diversity. Hewitt (1996) made a number of observations of genetic consequences arising from these processes. Firstly, he predicted that glacial refugia would contain populations with high genetic diversity, and that populations from different refugia would be genetically differentiated. Hence when the climate warmed, species would spread North rapidly; in doing so, areas may be colonized by individuals from one refugium or from several,

during which genomes may mix to various degrees. Populations geographically isolated in Southern refugia are likely to have undergone genetic differentiation due to drift, thus giving rise to detectable admixture between these genotypes during expansion.

Secondly, long distance dispersants would expand rapidly with their alleles dominating leading populations as they increase exponentially. Advancing at the leading edge would involve numerous bottlenecks during which alleles are lost and homozygosity would increase. Indeed modeling of long distance dispersal shows that it results in spatial clustering of genotypes that can persist for hundreds of generations (Ibrahim et al. 1996).

Subsequent research has shown that under certain conditions similar patterns can also emerge through continuous range expansion, without invoking long-distance dispersal (Hallatschek et al. 2007). This phenomenon is centered on the concept of 'allele surfing' (Excoffier & Ray 2008), where strong drift in small range edge populations can lead to large allele frequency changes in newly colonized areas. Empirical studies in European tree species have generally supported this. A region wide analysis of 22 European tree species by Petit et al. (2003) found the most genetically diverse populations at intermediate latitudes. They attributed these 'melting pots' to the admixture between divergent lineages colonizing from separate refugia. Theoretically, these could be confused with unidentified cryptic refugia, complicating the interpretation of modern population structure. However, true cryptic refugia would be expected to show higher variation and private alleles compared to surrounding recolonized regions.

1.1.3 Life history traits and succession

Whilst the overall colonization history of temperate tree species is becoming increasingly clear, questions remain over how patterns contrast in species with different life history traits. This research will become increasingly important as future climate changes, for us to understand which species groups may be most vulnerable. A study by Bhagwat & Willis (2008) assessed

whether species with evidence of persistence at Northern latitudes differed in biogeographical traits. In contrast to the large-seeded trees that survived only in southern refugia, they found that wind-dispersed generalist trees with the ability to reproduce vegetatively were more likely to have persisted in northern refugia.

This is remarkably consistent with life history characteristics identified in successional species as facilitating faster migration, such as faster growth, greater seed production, seed set at a younger age and longer distance dispersal. Curiously, Petit et al. (2003) found evidence that founding events had the strongest effect in species with the lowest dispersal abilities such as *Carpinus* and *Corylus*. Thus early successional species (which are also the most likely to have persisted in Northern refugia) may be expected to maintain more diversity on expansion than slower species arriving later from Southern refugia. Thus these two patterns may easily be confounded.

1.1.4 Warm-stage refugia and receding edges

In contrast to larger tree species, the distribution of arctic/alpine plants and dwarf vegetation is poorly known. Indeed it has also been questioned whether the distribution of alpine plant species during the LGM should be considered 'refugial' (Birks & Willis 2008). It is possible that population size and genetic diversity was more extensive historically than at present. Thus modern-day distributions could be termed 'warm-stage refugia' (Bhagwat & Willis 2008; Gentili et al. 2015). Hampe & Petit (2005) highlight that rear edge populations at the low-latitude margins of species ranges, are disproportionately important for long-term conservation of genetic diversity, yet are comparatively less studied. In contrast to species distributions that are in equilibrium, those with recent climate-driven range shifts harbor much of the species' genetic diversity at range edges, particularly at the receding Southern limit. These have been termed 'trailing edges'.

1.2 Threats to the conservation of woody plants in Europe

The current European flora has survived substantial environmental change over the quaternary cycle, with the current genetic and species diversity and distribution becoming largely established during the Holocene, as described previously. As a consequence, extant species must have evolved a level of resilience to severe range shifts, population fragmentation, bottlenecks and climate change in order to persist today (Davis & Shaw 2001).

Despite this, modern environmental changes emerging from the anthropocene pose new threats and challenges for the persistence of many plant species in Europe (Anderson 2016). These include human induced climate change (Engler et al. 2011; Thuiller et al. 2006), novel species interactions (Day & Leather 1997; Rhymer & Simberloff 1996), land use change (Rounsevell et al. 2006), habitat fragmentation (Vranckx et al. 2012) and resource exploitation (Ryan et al. 2000). Some of these threats are comparable to natural pressures during recent ice ages, whilst others are novel. However, all are likely to be experienced over shorter timescales.

Europe's forests are likely to be a bellwether for the fate of other temperate forests, which account for much of the world's woody biomass (Ehrlich 1996). Thus accurately interpreting the effects of these processes on species genetic diversity is essential for effective management. The recent State of Europe's Forests report (FAO 2015) highlights significant gaps in geographical representation of areas managed for in situ genetic conservation. These areas currently stand at 0.5 million hectares of a total 215 million hectares forest cover, whilst only 30 million hectares of forest have some form of protection.

Here some of the key threats are reviewed, together with an outline of the approaches with which to study them in woody tree species.

1.2.1 Historical population decline

Forest is the climax ecosystem in most parts of Europe under current climatic conditions, and at its maximum extent after the LGM, would have covered around 80% of the continent (FAO

2014). Following several centuries of decline, in the 19th century forest cover was reduced to less than a quarter of the Europe's land area. Heavy reforestation over the past 150 years has result in forests covering 33% of the total land area and increasing. However, as a result many forests are relatively young and only semi-natural. For example, more than half are under 80 years old and even-aged.

Drivers of this overall decline can broadly be encompassed as unsustainable exploitation of forest resources and shift in land-use to managed agriculture. Many areas of natural forest have also been converted to plantations, which, whilst maintaining populations of some tree species, often select for certain phenotypes or include non-native species (Paques 2013). As a consequence, the effective population size of many European tree species is likely to be less than during previous interglacials, and some species may have experienced a mild recent bottleneck. It is also likely that as species become rare or locally extinct, unique genetic diversity from parts of a species' range is irreversibly lost (Koskela et al. 2013).

1.2.2 Fragmentation and gene flow

Deforestation and land-use change in Europe has been highly heterogeneous. Whilst forests were reduced to 25% of their historical extent region-wide, current forest cover ranges from 10% to 68% in continental European countries with substantially greater cover at higher latitudes (FAO 2015). Surviving forests occur in a highly fragmented landscape with poor connectivity (Estreguil et al. 2013). In the EU 40% of forests are within 100m of other land use classes, whilst 70% is classified as poorly connected. Fragmentation has also been hypothesized to be most prevalent at species range edges (Hampe & Petit 2005), which are likely to hold important genetic diversity.

Population genetics theory predicts several consequences of fragmentation on the long-term viability of species (Kettle & Koh 2014). Primarily, genetic diversity is predicted to be eroded as a result of restricted gene flow due to population isolation and increased genetic drift in small

populations. A consequence of reduced gene flow is limited dispersal of novel advantageous mutations between fragments, limiting local adaptation of populations and the adaptability of the species as a whole. Genetic drift is the change in allele frequencies due to random sampling of gametes from the parental population. Whilst in large populations the effects of drift are small, in small populations drift can result in rapid differentiation, the loss of rare alleles and fitness impacts.

Furthermore, by reducing local effective population sizes, habitat fragmentation has also been shown to increase inbreeding (Ellstrand & Elam 1993). Inbreeding is the mating of related individuals resulting in increased homozygosity. Drift and inbreeding can allow deleterious allele to increase in frequency, or even reach fixations, reducing the fitness of the population (Leonardi et al. 2012). As a result of these processes, fragmented populations are likely to become increasingly differentiated. An anticipated outcome across the species range is higher genetic differentiation among populations, combined with lower levels of within-population variation (Bacles & Jump 2011). Fragmentation may also negatively affect mating systems and demographic structure, which in turn influences seed production and progeny viability (Finger et al. 2012). The finer the fragmentation scale and the less tolerant the life history traits of the species (such as a lower propensity for long distance gamete dispersal), the greater these effects on genetic structure and species ecology are likely to be.

Equilibrium is when forces affecting genetic variation are balanced and metrics such as genetic diversity or differentiation are not changing from one generation to the next. When considering the ecological genomics of European tree species, it is challenging to account for the fact that many populations are not at equilibrium, but rather are in a transient state. Many factors influence the time lags to reach equilibrium. Genetic time lags can bias ascertainment of population genetic consequences from fragmentation and climate change. Consequently, recent assessment of genetic structure in European tree species may more strongly represent past landscapes. This makes it difficult to unravel how contemporary anthropogenic activity

such as forest fragmentation or changes in population structure, size and connectivity are reflected in genetic metrics (Epps & Keyghobadi 2015). Furthermore, true values may be difficult to infer because genetic parameters of interest typically do not reach equilibrium in a linear fashion (Whitlock 1992).

1.2.3 Climate change

Annual temperatures in Europe are predicted to rise by 0.1-0.4°C per decade through this century (Pachauri & Meyer 2014), thus a rise exceeding 2°C is considered likely by the year 2100. Annual average land temperatures in Europe are projected to increase by more than the global average with precipitation increasing in Northern Europe and decreasing in the South (Eu et al. 2015). The intensity and frequency of extreme weather events is also anticipated to increase (IPCC 2014a), resulting in greater variability in conditions for Europe's tree species. In addition to direct effects (temperature, rainfall etc.), evidence is also mounting for indirect negative consequences arising from climate change including increased pest, disease and fire incidence (Alfaro et al. 2014).

Empirical observations and species distribution modeling predicts significant poleward range shifts of European tree species in response to climate change (Parmesan 2006). However the dispersal required to realize these shifts exceed current maximum estimates of post-glacial migration rates (Aitken et al. 2008). Thus a combination of plasticity and evolutionary adaptation, as well as range shifts, will be required for tree populations to successfully track climate change (Alberto et al. 2013).

Trees, being sessile and long-lived, have generally evolved a high level of tolerance to a range of environmental conditions, termed plasticity. Nicotra *et al.* (2010) showed that plasticity in seed longevity, phenology and other traits have been documented in response to elevated CO^2. Higher levels of genetic variation in populations may result in greater range of plastic responses in a given population (Falk & Holsinger 1991). Thus we may conclude that plasticity

can provide a buffer in the context of rapid climate change, and assist adaptation. Greater plasticity does however mean weaker selection pressure which may interact with the rate of evolutionary adaptation (Via & Lande 1985). Furthermore, an adaptive response is predicted to improve when populations are large, have high levels of genetic variation and there is suitable habitat for the establishment of new and better adapted genotypes (Alberto et al. 2013). These responses are also predicted to occur faster where selection is stronger. Thus loss of genetic diversity from reduced population size, forest fragmentation and isolation may impede a species ability to adapt to changing environmental conditions.

Gene flow can have complex and sometimes contradictory effects on populations, both constraining local adaptation and enhancing selection in certain situations (Kremer et al. 2012). In the first scenario high levels of gene flow homogenises allele frequencies, as such, population differentiation is reduced even when a species occurs across an environmental gradient. Across a well connected distribution, the center of a species range should contain the largest fittest populations, with smaller less well adapted populations at the periphery close to the niche/range limit (Eckert et al. 2008). Thus gene flow outwards to the range edge may swamp and prevent local adaptation (Chevin & Lande 2011). By contrast, reduced gene flow may facilitate local adaptation of genotypes to closely match environmental conditions, at the risk of reduced effective population size and increased genetic drift. Overall, as the suitable bioclimatic envelopes for each species shifts, this is likely to result in significant reorganization of species genetic diversity. If species are unable to track this shift, tolerate conditions or adapt, this will result in reduced fitness or extinction.

1.2.4 Introgression and hybridisation

During previous climatic oscillations there is evidence for hybridization and introgression in several European tree species (*Abies, Liriodendron, Picea, Quercus*; Comes & Kadereit, 1998). Indeed many European tree species have congeneric infertile hybrids in some part of their range. Levin and authors (1996) showed that hybridization can contribute to the decline of

plant species through demographic swamping and genetic assimilation. In severe scenarios hybridization may even bring about the extinction of a species (Rhymer & Simberloff 1996). Conversely the benefits of hybridization, such as adaptive introgression and increased genetic diversity, are increasingly being recognized (Becker et al. 2013; Hamilton & Miller 2015; Brennan et al. 2014; Mallet 2005). As is its role historically as a key driver of speciation (Thomas 2015). Thus the opportunity or threat posed from introgression and hybridization to Europe's tree species must be carefully assessed.

Climate change, fragmentation and the anthropogenic dispersal of plants around the world is likely to bring about novel species assemblages and interactions. Whilst past range expansions and contractions were likely to produce parapatric distributions between species, under fragmentation and species invasion many plants display a 'honeycombed' distribution increasing the surface area of contact and accelerating any eventual outcomes of that interaction (Levin et al. 1996). Indeed there is much evidence that invasion success and spread of novel hybrids is increased by disturbance and fragmentation of habitats (Vila et al. 2000; Petit et al. 2004). Some alternative scenarios particularly in the context of range shifts, climate change and montane species are given in Figure 1.3 adapted from Gomez *et al*. (2015).

Hybridization is widely predicted to result in a loss of genetic diversity and impact on locally adapted populations. Research on extinction risks from hybridization suggests that it is perhaps the most rapidly acting genetic threat to endangered species (Wolf et al. 2001), with extinction occurring in as few as five generations. Hybridization can contribute to species decline in two main ways (Ellstrand & Elam 1993). Firstly, if the species has reduced fitness and/or frequency relative to another hybridizing species then the less fit or numerically inferior taxon may decline below the rate of replacement (demographic swamping). Secondly, if hybrids are fertile and of equal or greater fitness they may displace one or both of the parental species (genetic assimilation). The authors find that numerous factors, ranging from pollen-tube growth rate to selfing rate and number of patches, influence the outcome of

hybridization, but small population size in one of the taxa makes a detrimental outcome likely (Petit 2004; Beatty et al. 2010; Buggs 2007).

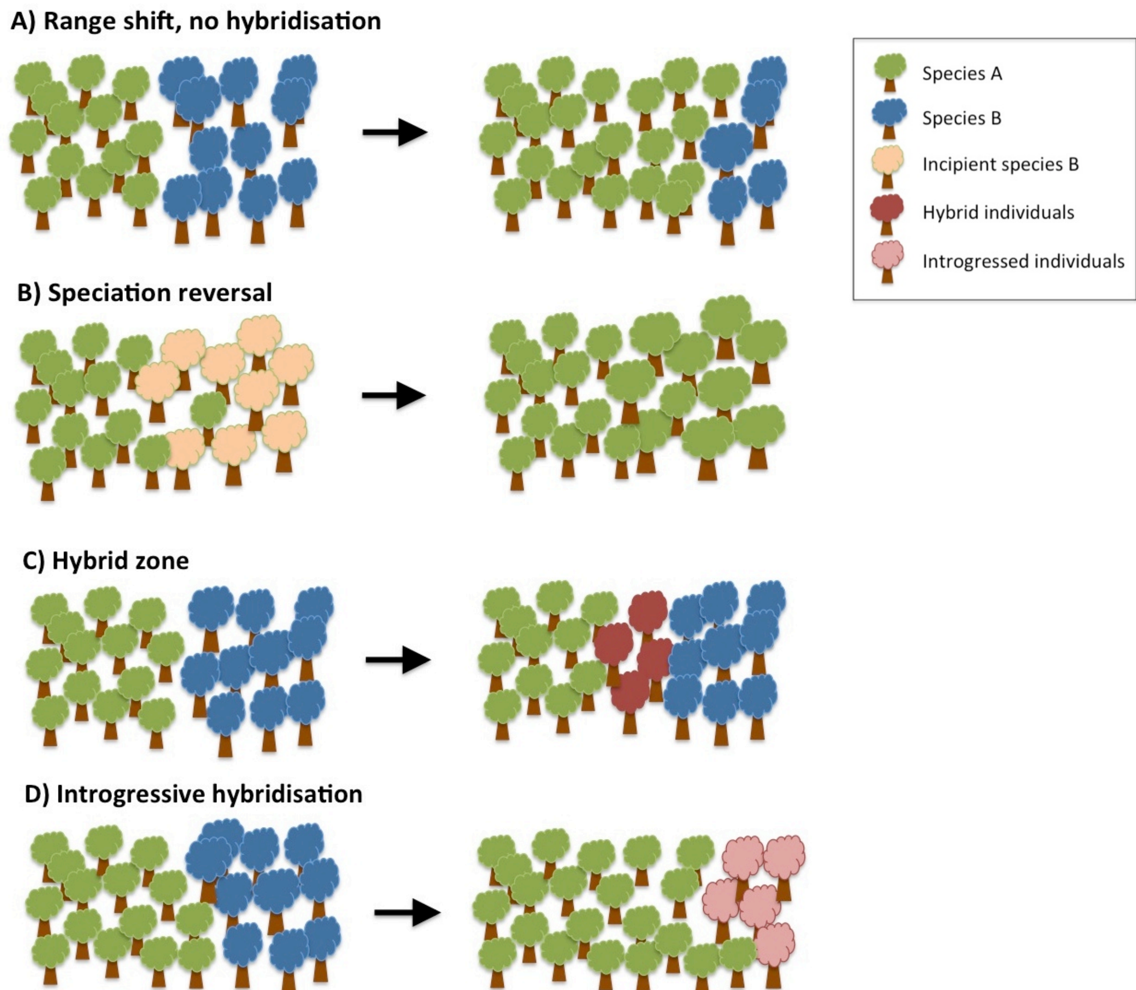


Figure 1.3 Consequences of contact between species ranges. A) No hybridisation occurs, perhaps due to reproductive or other barriers. The result is decline of the less competitive species. B) Speciation reversal where there is an insufficient reproductive barrier between species. This results in a single, genetically homogeneous species. C) Non-introgressive hybridisation where hybrids occur without extensive backcrossing, perhaps due to the intermediate genotypes being less fit or infertile. The result is a hybrid zone, that may move over time. D) Introgressive hybridisation occurs where there is extensive gene flow between two species and introgressed individuals. Where one species is significantly more abundant, the other may become extinct, but introgression may also occure from the declining species into the advancing species.

Interactions at range edges are likely to be particularly important in the context of climate change. An assessment of hybridization at range limits in Canadian *Pyrola* populations found that extensive unidirectional hybridization may lead to the extinction of peripheral populations. Importantly this could compromise the ability of the species to respond to climate change via habitat tracking (Beatty et al. 2010).

The potential for adaptive introgression to increase species' ability to respond to climate change has also been explored (Hamilton & Miller 2015). In some scenarios, interspecific gene flow could augment limited adaptive potential where there are low levels of standing genetic variation. This is likely to be most beneficial in species that exhibit strong ecotypic differentiation (to avoid negative consequences discussed above) and where the hybrids display traits outside of the parental phenotypic range – termed transgressive segregation. However, in some cases hybrids may experience higher fitness resulting in rapid allele frequency shifts and local extinction of one or both parental species. This can occur quickly or it may require several generations of backcrossing to achieve a superior fitness of introgressed forms.

From a conservation perspective, the potential risks of introgression and hybridization appears to be commonly understated in the literature when compared to forest fragmentation, climate change and other pressures. It is important due to the rate at which genetic information can be irreversibly lost, however it is not necessarily detrimental in all cases (see Brennan et al., 2014 for an excellent discussion). Hybrids or introgressed populations may survive and become dominant if they display increased fitness – however under rapid environmental change those genotypes may not retain increased fitness long term and will subsequently poses reduced standing genetic variation, limiting future adaptive potential (Hoffmann et al. 2011).

1.2.5 Pests and diseases

Until recently, even considering glacial expansion and contraction, geography and climate ensured a relatively stable distribution of plant species over timescales of centuries. However, human activity has recently accelerated the redistribution of species around the world (Santini et al. 2013). In a study of 123 pathogens, Santini and authors (2013) found that the number of invasive forest pathogens (IFPs) introduced to Europe has increased exponentially in the past 200 years. The highest numbers were reported in central-southern countries such as Italy and France. Similarly the Program for Monitoring Emerging Diseases has for example shown a 13-fold increase of plant infecting fungi reported from 1995 to 2010 (Fisher et al. 2012). Many of these are likely to have evolved outside of the range of European tree species. Indeed only 28% of IFPs identified since 1800 were classified as of European origin (Santini et al. 2013). Species may be commensal or mildly pathogenic in their original range or on co-evolved hosts, yet become highly pathogenic elsewhere where new hosts have no evolved resistance mechanism (Boyd et al. 2013).

Where newly introduced pathogens are harmful, affected species are at risk of depletion of genetic variation as a result of reduced effective population size and loss of locally adapted populations. However it is worth noting that even severely affected species (e.g. White elm, *Ulmus laevis*) still remain in the European landscape and have escaped extinction, perhaps due to high levels of diversity in woody plant species (Collin 2003). Boyd *et al*. (2013) do note that their review on tree pests excludes large mammalian herbivores such as deer, yet their role on tree populations potentially has a much larger effect than most pests and diseases

Pests are unique among the threats posed to European trees in that eradication seems largely impossible, with prevention by early detection the most prudent strategy for the future. Thus reducing negative pressures from other threats such as climate change and forest fragmentation seems to be one of the few available options to manage IFPs.

1.2.6 Research Challenges

It is clear that Europe's tree populations are unlikely to experience any of these threats in isolation. Rather, multiple drivers occurring in synergy are likely to place additional stress on species. Thus current knowledge gaps include:

1. What is the current distribution of genetic diversity in montane species, and is it consistent with more extensively studied lowland trees?
2. Which pressures pose the greatest threat to montane species and their persistence in the immediate future?
3. By comparing neutral and selective markers, is there any evidence for local adaptation in montane species, and is the potential for adaptive introgression from lowland species likely to be advantageous under climate change?
4. How can we synthesize the inferences from ecological genomics to conserve *in situ* genetic diversity?

1.3 Genus *Betula* as a study system

The Betulaceae comprises six genera and likely evolved in central China, the center of species diversity for this family (N. Wang et al. 2016; Ashburner & McAllister 2013). Within this family, species of the genus *Betula* provide an ideal system in which to elucidate the genomic processes underlying the diverse threats to European forest and montane trees. In Northern Europe there are three naturally occurring species from the genus; *B. pubescens*, *B. pendula* and *B. nana* (Ashburner & McAllister 2013). All are pioneer species that would have been amongst the first to colonize Northward as the glaciers retreated after the LGM (Feurdean et al. 2013). Furthermore, as wind-pollinated trees, birch pollen is well represented in sediment samples from the Holocene, allowing reconstruction of forest history (Huntley & Birks 1983; Birks 1968; Caseldine 2001).

1.3.1 Ecology of *Betula* species in Europe

Betula pubescens and *B. pendula* are deciduous trees growing to 10-20m under ideal conditions. Due to strong phenotypic similarities they have historically been treated as conspecific (*B. alba*) by some authors (Tuley 1973). They have a broad overlapping range from the Mediterranean, north to Lapland. Both are early colonizers, with *B. pubescens* more common on poorly drained sites such as peat bogs and in areas with higher rainfall. *B. pendula* is more common on dryer sites and tends to favour lowland areas. Some texts also discuss Mountain Birch *B. tortuosa* (Truong et al. 2007), however most researchers consider this a subspecies or northern ecotype of *Betula pubescens (*N. Wang, *pers comm)*.

Although moderately closely related (*B. pubescens* may be the progenitor of *B. nana* and another unidentified *Betula* species, Wang et al. 2016), *B. nana* is a subarctic, circumpolar dwarf tree occurring at and above the tree line and on sub-arctic tundra, and differs markedly in phenotype (Figure 1.4). *Betula nana* frequently displays prostrate growth and is characterized by copper-red bark and small rounded leaves with a bluntly toothed margin. It is

monoecious and wind-pollinated, with wind dispersed winged nutlets and is thought to be self-incompatible, consistent with many other *Betula* species. *B. nana* is intolerant of light limitation, and severe waterlogging, though it favors peat bog habitats (Ejankowski 2008; Ejankowski 2010). Across its range it comprises two largely allopatric subspecies, *ssp. nana* (Suk.) Hult. From Greenland, across Europe to Western Asia and *ssp. exilis* (Suk.) Hult. In North America and Eastern Asia, *ssp*. *exilis* is also more recently treated as *B. glandulosa* (Ashburner & McAllister 2013).

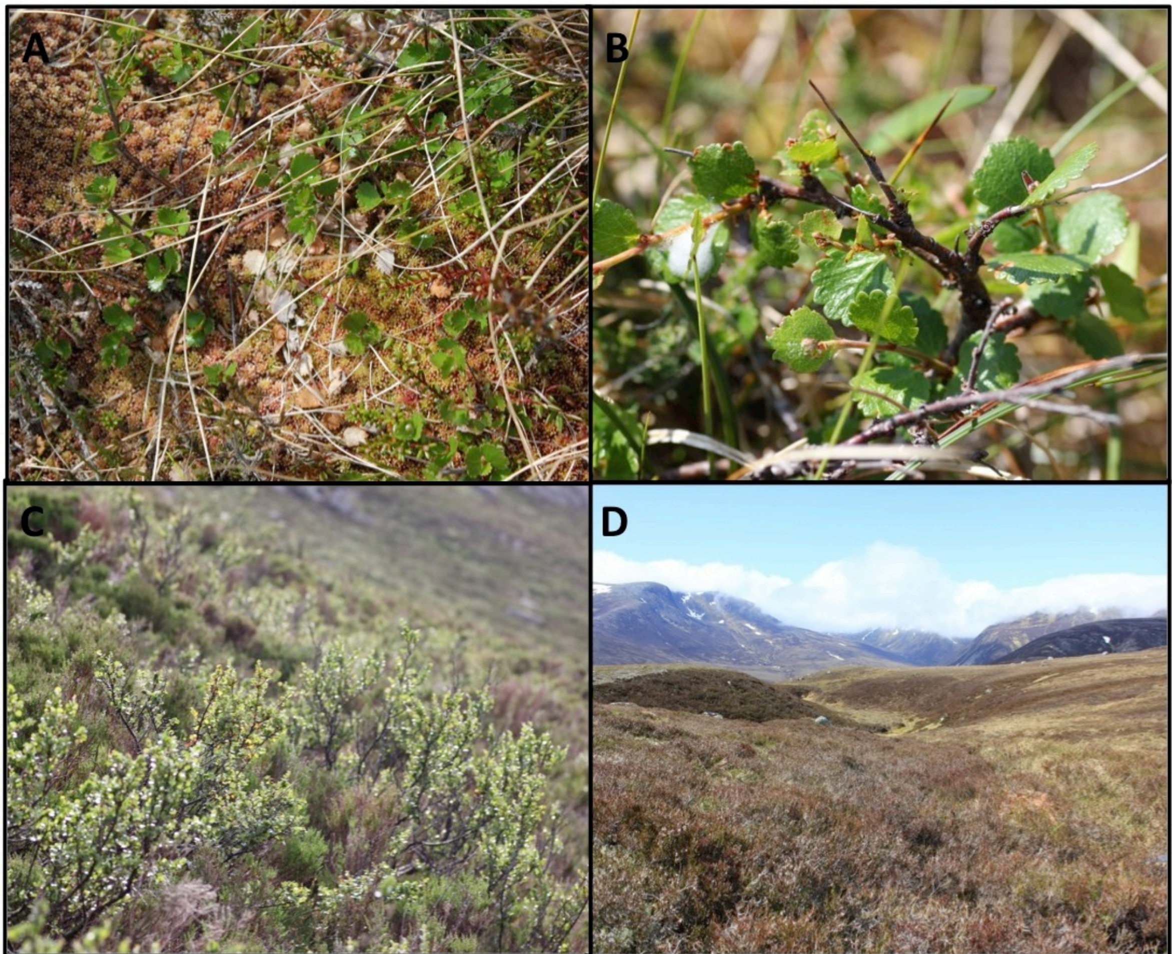


Figure 1.4 *Betula nana* individuals from across the Scottish highlands. A) *B. nana* shoots emerging from sphagnum moss. B) A *B. nana* plant showing signs of heavy grazing and stunted growth. C) A large, healthy *B. nana* population growing on a well drained 40 degree slope at Loch Muick, Scotland. D) Typical *B. nana* habitat, near the river Avon, Cairngorms, Scotland.

1.3.2 *Betula nana* and montane scrub

In Scotland *B. nana* forms a key component of montane scrub habitat (Gilbert & Di Cosmo 2009), together with several rare *Salix, Vaccinium* and *Juniperus communis*. Montane scrub is the zone between woodland and higher altitude heath where plant growth is restricted and stunted (see Figure 1.4). The result is a natural transition between habitats, rather than an abrupt change where commercial forestry no longer becomes viable (Figure 1.5). A challenge for the restoration of montane scrub habitat is that the natural treeline varies locally with climate, aspect and exposure, thus it remains difficult to predict where the original vegetation is no longer present.

Whilst in many regions *B. nana* readily produces seed and a seed bank, there is evidence that sexual production is suppressed where conditions are below optimum in relict populations, for example in Svalbard (Alsos et al. 2003) and potentially in Southern Scotland (*pers. comm.* Highland Birchwoods). Ripe seeds have been observed in Greenland, whilst a germination rate of 52% was reported from a low Arctic site in Alaska and only sterile *B. nana* was observed on Novaya Zemlya (Soyrinki 1939). Similarly, at the northern limit of its distribution *B. glandulosa* shown almost complete absence of sexual reproduction (<0.5% viable) thus reproduces vegetatively (Weis & Hermanutz 1993).

Vegetative reproduction is thought to be common in established populations, and this can make reliable identification of individuals difficult (Alsos et al. 2002). Researchers found that plants at their range limit produced both fewer pollen catkins, less pollen per catkin and a lower proportion of the pollen is viable (Weis & Hermanutz 1993). Thus it is hypothesized that the lack of sexual reproduction is partly due to pollen limitation. Indeed Handel (1985) found that larger clones set fewer seeds due to increased incidence of geitonogamous pollination. Furthermore in this species phenology does not decrease geitonogamous pollination events as the male and female catkins mature almost simultaneously, possibly due to the short growing season across much of its range.

Finally, *B. nana* represents an unusual case study in the context of postglacial expansion. Of all species that may have survived in Northern refugia, if they existed, *B. nana* is highly likely to be one on account of longevity, prostrate growth, clonal reproduction, adaptations to annual snow cover and current distribution on tundra habitats in Siberia and Alaska (Henry & Molau 1997; Ejankowski 2008). Similarly, pollen records and vegetation reconstructions, reported above, place *B. nana* alongside advancing tree birches and other pioneer species with rapid northward migration rates. However demographic and life history traits for *B. nana* appear remarkably inconsistent with characteristics that facilitate faster migration such as fast growth and higher seed production. Whilst these traits may hold for larger *Betula* species, these traits do not appear to be true of *Betula nana*.

1.3.3 Drivers of *Betula nana* decline in the UK and Europe

Whilst *B. pubescens* and *B. pendula* are relatively widespread and abundant, by contrast *B. nana* is common only at higher latitudes and elevations across northern Scandinavia. It currently persists in a handful of relict sites in the Alps and Carpathians (Kruzelnicki & Fabiszewski 2001), and just three small lowland populations in Poland (Drzymulska 2014). In the UK *B. nana* is listed as nationally scarce (Stewart et al. 1994) and is subject to conservation programmes led by Trees for Life, Highland Birchwoods, Scottish Natural Heritage and the Montane Scrub Action Group.

Possible drivers underlying the decline of *B. nana* align closely with threats to European forest species in general. Firstly, the UK appears to be at a retreating Southern bioclimatic range limit, and is composed of numerous fragmented and isolated populations of varying size, whilst local extinction has occurred in some areas (Hutchinson 1966; Aston 1984). Second, as a montane species, it is also likely to be threatened by climate change with habitat loss predicted to be greater for species at higher elevations (Engler et al. 2011), whilst moorland burning and grazing regimes may also be detrimental (Holl & Smith 2007). Finally, *B. nana* is also known to hybridize extensively with other *Betula* species elsewhere in its range (Palme et

al. 2004a; Thórsson et al. 2007; Elkington 1968; Karlsdóttir et al. 2009; Anamthawat-Jónsslon & Tomasson 1999), with some reports of hybrids also occurring in the UK (Aston 1984). By means of comparison, *B. pubescens* and *B. nana* also occur extensively across Northern Lapland, where human disturbance is limited and population abundance and distribution should be in a semi-natural state and well within bioclimatic range limits (Bergman et al. 2005; Makela 1998).

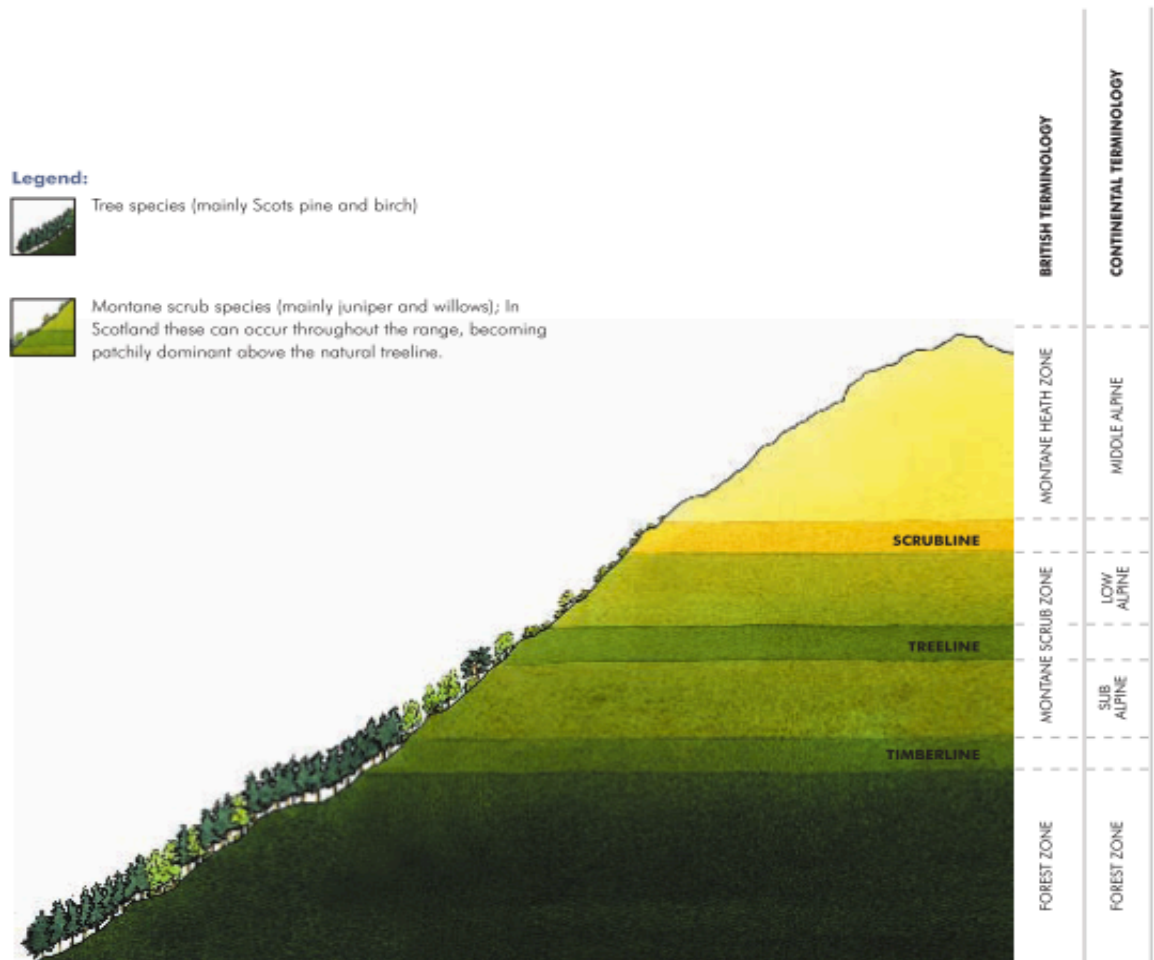


Figure 1.5 Theoretical zonation of tree species and habitats across Scottish Mountains. Reproduced with permission from (Scott 2000).

1.3.4 Hybridization and polyploidy in *Betula*

Natural hybridization is widely reported from within the family Betulaceae and genus *Betula* (Makela 1998; Karlsdóttir et al. 2009; Anamthawat-Jónsson & Thórsson 2003; Thórsson et al. 2001; Kallio et al. 1983). Of the species discussed in this thesis, *B. nana* and *B. pendula* are

diploid, whilst *B. pubescens* is tetraploid. Ploidy differences are often considered to be an insurmountable barrier to geneflow as offspring are generally sterile due to uneven chromosome segregation and variation in gene dosage (Griffiths et al. 2000). Despite this, an increasing number of studies are suggesting that successful hybridization across ploidy levels may occure, however infrequent (Clark et al. 2015; Wang et al. 2014; Anamthawat-Jónsson & Thórsson 2003; Bohlen et al. 2016). Natural hybridization in birches is likely to lead to gene flow between species through backcrossing of hybrid individuals with the parental species. As such, it has been regarded as one of the most important pathways of evolution in birches (Kallio et al. 1983; Elkington 1968).

Extensive hybridization has been reported between *B. nana* and *B. pubescens* in wild populations. Research in Iceland found that around 10% of birch plants collected randomly were triploid (2n = 42) hybrids, confirmed by genomic *in situ* hybridization (Anamthawat-Jónsson & Thórsson 2003). Tree birch hybridization appears to be limited by a temperature dependent incompatibility system (Hagman 1971), differences in flowering time and niche separation (Kallio et al. 1983). However at Northern range limits the incompatibility system is less effective, thus introgressive hybridization becomes more common (Makela 1998; Palme et al. 2004b). Hybridization is further permitted by short growing seasons which favor greater synchronicity in flowering (Kallio, P. et al. 1983). In the field, due to selection against recombinant intermediate phenotypes, morphological measurements may underestimate the true proportion of hybrids, thus molecular markers are advantageous to determine actual hybrid frequencies (Rieseberg et al. 1999).

1.3.5 Current research on European Birches

Recent research has been conducted by Eidesen et al. (2015) to disentangle the phylogeographic history of European *Betula* species during and since the last glaciation. A clustering analysis across the species range identified five groups along an east-west gradient which the authors attribute to a more ancient differentiation process. Relict populations on

Svalbard were genetically most similar to those of West Siberian origin, supporting long distance dispersal. Interestingly the authors found incongruence between plastid DNA and AFLP variation in *B. pubescens*, suggesting plastid haplotype sharing emerging from chloroplast capture by *B. pubescens* as it progressed North. Contrary to expectation, the authors also found an increase in genetic diversity northwards, though this may well be attributed to lineage admixture and loss of genetic diversity at the rear range edge. The study is however limited by the number of samples sequenced at plastid DNA regions, and the overall number of markers.

Research on *B. pubescens* has found weak isolation by distance and little evidence for population differentiation. This possibly reflects phenotypic and demographic differences. Absence of large-scale structure suggests high levels of current gene flow through both seed and pollen (Eidesen et al. 2013; Thórsson et al. 2007; Truong et al. 2007). It is plausible that across Europe the outcompeting *B. pubescens* was continuously expanding into the range occupied by *B. nana*. Pollen and macrofossil data support this (Bergman et al. 2005), and the complete role of this process in the evolution of both species remains to be fully elucidated.

On a broad scale, analysis of 17 widespread arctic-alpine plants attempted to assess whether phylogeographic patterns were species-specific or congruent despite different life history and dispersal characteristics. The authors found distinct trends that could be explained by glacial oscillations and known geographic barriers such as oceans, ice caps and mountain ranges, consistent with predictions (Eidesen et al. 2013). Evidence was also found that Northern Europe was colonized from the East as well as the South, supporting previous studies. However the population structure of *Betula* species on a regional scale and the impacts of population decline and fragmentation remain unexplored.

1.4 Utilizing ecological genomics to assess threats

In this study we use newly emerging genomic techniques together with phenotypic and ecological climate data to understand the drivers of decline in *B. nana*. This species represents a unique combination of dispersal and life history traits, largely absent from the current literature, and together with related birch species provides an ideal system in which to elucidate threats posed to montane woody tree species in Europe. Here I outline the approaches that are undertaken in subsequent chapters to address these challenges.

1.4.1. Range-wide sampling for predicting current and future distributions

The first phase of this study aims to establish the current distribution and condition of *B. nana* populations across the UK. This is supported by a combination of historical and contemporary records from partner organization, and fieldwork surveys. For subsequent analysis, this enabled sampling across the entire extant range in the UK, encompassing range limits and the Southern receding edge. For comparison we also sampled across a comparative geographic scale in Scandinavia where the impact of anthropogenic drivers is substantially lower. Preliminary observations have suggested that UK *B. nana* plants are smaller, less productive than in Scandinavia, potentially as a result of browsing, periodic burning and climate (Montane Woodland Survey, 2009). Furthermore, herbivore enclosures have been proposed for remediation to encourage population regeneration (pers. comm. Trees for Life). Thus we conduct a region wide study to assess variation in phenotypic traits across the UK.

Ecological niche modeling (ENM) and genotype-environment correlations have provided fresh insights in to the distribution, tolerance and adaptive potential of species (Rellstab et al. 2015; Guisan et al. 2013). For example, ecological metrics may predict genetic connectivity better than simple geographic distance (Hokit et al. 2010). Collation of historical and contemporary records, together with high resolution environmental data allows the prediction of species distribution, which is particularly useful for cryptic montane species. Here we use ENMs to

predict the current and future distribution of *B. nana*. In addition, we attempt to further validate ENMs using correlation with phenotypic fitness measurements, with the aim of improving their predictive performance.

1.4.2 Combing inferences from microsatellite and SNP markers to establish historical distribution

A key obstacle to understanding how landscape scale processes influence gene flow and genetic structure is lag time in population genetic metrics. In a guide to landscape genetic inference, Epps et al. (2015) recommend using historical landscape distribution data in unison with genetic markers, as this may inform the investigator about different time periods. Furthermore, they identify comparative analysis of lag times in neutral and selective markers as a key area of research.

Different markers such as AFLPs, microsatellites and SNPs are subject to different mutation rates, with higher mutation rate markers such as microsatellites likely to approach equilibrium at a faster rate (Beaumont & Nichols 1996). By using a multiple independent markers, such as microsatellites and SNPs, it maybe possible to exploit this difference gain greater insights into the demographic history of *B. nana*. Furthermore, in the context of fragmentation and population history in European birch species, equilibrium is likely to be approached much more rapidly in small populations of low effective size. Thus an appropriate strategy must include a representative sample of extant populations across a range of population sizes. Here we explore the relative contribution of different mutation rate markers and investigate whether combining markers improves our inference of the demographic history of dwarf birch.

Conversely, the development of large numbers of microsatellite loci in addition to next generation sequencing approaches may be prohibitively time consuming (Hodel et al. 2016). Thus we also develop a novel method to derive higher mutation rate markers from short sequencing reads to enable this approach to be more widely applied.

1.4.3 Detecting hybridization and introgression

The propensity of species to adapt to pressures such as climate change via adaptive introgression has received less attention than more conventional responses including migration and phenotypic plasticity. Interspecific gene flow may provide an alternative mechanism to escape the limited adaptive potential from low standing genetic variation and mutations alone (Hamilton & Miller 2015). Whilst hybridization is generally regarded as harmful in a conservation context (Gomez et al. 2015; though see Becker et al. 2013), here we use a range of approaches to assess individuals' ancestry composition, explore evidence for adaptive introgression in *Betula* species and discuss the potential consequences.

Whilst adaptation to climate change is the most obvious pressure that may drive adaptive introgression, it is also plausible that species may have adaptations that increase fitness in response to fragmentation, grazing, disease and other threats (Kawecki & Ebert 2004; Taylor et al. 2015; Hamilton & Miller 2015). Conversely, this process could result in extirpation of a parent species, outbreeding depression and disruption of locally adapted genotypes. From a gene-centric perspective (Petit 2004), hybrids may be considered as evolutionary repositories of adaptive loci from both parents, and in extreme cases might offer a capacity for genes to persist despite extinction of a parental species (Rius & Darling 2014). In *B. nana*, hybridization is known to occur both in the UK (Wang et al. 2014) and elsewhere in the range (Anamthawat-Jónsson & Thórsson 2003), but the extent to which this influences the species evolution is unknown.

1.4.4 Detecting signatures of local adaptation

A crucial question for European tree species is whether populations are capable of tolerating, migrating or adapting to novel pressures (Meier et al. 2012; Savolainen et al. 2007). Detecting signatures of adaptation in species genomes can indicate past evolution and allow inferences to be made about future adaptive potential. There are many genomic patterns that could indicate selection, including regions of reduced variation, skewed site frequency spectra,

linkage disequilibrium, loci-environment correlations and more broadly, elevated rates of divergence between populations or species (Jensen & Foll 2016). Other processes may confound this signal, producing genomic patterns resembling selection such as changes in population size or structure and background selection.

Many approaches utilize inter-population measures such as F_{ST} to reveal loci where alternative alleles have been fixed in different populations, resulting in greater difference in allele frequency that expected under neutral evolution (Beaumont & Nichols, 1996). These produce few false positives, but fail to detect selection outliers when selection is not strong (Pardo-Diaz et al. 2015). Despite numerous loci identified across genomic scans such as these, the number that have been linked to phenotypic or fitness traits is limited (see Hohenlohe et al., 2010; Suarez-Gonzalez et al., 2016).

More recent methods detect selection based on correlations of genetic data with environmental changes (Frichot et al. 2015; Forester et al. 2015; Rolland et al. 2015; Joost et al. 2013; Joost et al. 2007). Thus here we combine range-wide sampling encompassing all currently experienced climatic conditions with genome wide SNP markers to improve our ability to detect loci under selection, and associate these with environmental variables. Furthermore, we develop a metric to assess the degree to which allele frequencies deviate from the proposed optimum under a given set of environmental conditions. We explore the potential for this metric to inform conservation management of populations.

1.4.5 Planning for optimal conservation of genetic variation

As conservation science evolves, management practices and evolutionary theory must combine (Eizaguirre & Baltazar-Soares 2014). This research has been conducted in collaboration with several conservation organisations in Scotland, including CASE partners Trees for Life and Highland Birchwoods. A key outcome to support partner organisations is ecological and genetic evidence informing ongoing conservation and restoration of *B. nana* in

the Scottish Highlands and Northern England. Furthermore our research helps managers to prioritise the conservation of genetic variation that increases the potential for evolutionary adaptation.

B. nana populations are widely considered to be genetically depauperate in Scotland compared to other montane species that may have larger populations and greater dispersal ability (pers. comm. Trees for Life, Highland Birchwoods), on account of their small size and geographic isolation. Here we will use a systematic conservation planning (SCP) approach in which a minimum set of sites is recommended that maximize genetic diversity. Whilst traditionally used at the species level (Asmyhr et al. 2014; Brooks et al. 2006), more recently genetic diversity metrics have been incorporated to guide planning (Diniz-Filho et al. 2012; Vinceti et al. 2013).

A meta-analysis by Frankham (2015) found that outcrossing of inbred populations screened as having low risk of outbreeding depression, resulted in beneficial effects in 92.9% of 156 cases. Finally we discuss whether genetic rescue is likely to be beneficial or necessary for *B. nana* in Scotland and the potential applications in other European tree species. Future conservation strategies could consider inter- or intra- specific genetic rescue – augmented or artificial gene flow from other populations, or even species. Here we attempt to unify phenotypic fitness measurements, species distribution models, neutral and adaptive genetic diversity and the potential for assisted geneflow to give an overview of possible conservation strategies for Dwarf birch.

Chapter 2: Evaluating the distribution and fitness of *B. nana* under current and future climate

Summary

Historic climate change and shifting land use has shaped the assemblage and distribution of species across the UK. In the future, the increasing rate of anthropogenic changing climate is likely to result in greater disequilibrium between species distributions and environmental conditions. As a result, species distribution models are increasingly used to predict future habitat suitability and inform conservation management. However the relationship between habitat suitability and population fitness has rarely been tested. In this study we conduct the first detailed UK wide assessment of the past and present distribution of *Betula nana*, a nationally scarce and range restricted montane plant. We show that future emission scenarios are likely to result in substantial further range contraction across the UK and we identify regions that may act as climate refugia. We show that *B. nana* plants in the UK are significantly shorter and subject to greater browsing pressure than those in Scandinavia, but produce significantly more catkins, even after correcting for plant size. Catkin production and seed germination rate are also highly variable among populations. Finally we provide amongst the first empirical evidence that species distribution model derived habitat suitability values are significantly correlated with reproductive output, a proxy for fitness, lending support to their continued use in conservation management.

2.1 Introduction

Plant distributions correlate with abiotic conditions across a range of taxa and geographic scales (Stephenson 1990; Pfeifer-Meister et al. 2013; Guisan & Zimmermann 2000; Parmesan 2006). By identifying the subset of environmental conditions that characterize a species' niche, modeling approaches have been developed to predict current distributions and project future range shifts (Phillips et al. 2006; Pearson et al. 2007; Elith & Leathwick 2009). As a result, species distribution models (SDMs) have become a widely applied tool in biogeography and ecology (Pearson et al. 2007; Elith & Leathwick 2009; Guisan et al. 2013).

An implicit assumption of SDMs is that populations with a high predicted habitat suitability index (HSI), are likely to exhibit higher fitness than those occurring at sites with low habitat suitability (Guisan & Thuiller 2005). However, to date few studies have empirically tested this by relating model derived HSI with *in situ* metrics such as catkin production or seed germination rate that may indicate the likelihood of population persistence, regeneration or dispersal. This is pertinent because many species ranges appear to be shifting northwards and to higher elevations (Parmesan 2006; Evangelista et al. 2016) whilst changes in the physiology and phenology of plants attributed to climate change have already been detected (Thuiller et al. 2005; Hughes 2000; Parmesan & Yohe 2003). Thus despite numerous studies demonstrating the accuracy of SDMs for predicting species occurrence, they may fail to reflect the fact that relict populations could persist in climatically suboptimal conditions with little reproductive potential, thus yielding a false indication of tolerance (Springate & Kover 2014; Stanton-Geddes et al. 2012; Bjorkman et al. 2016).

Of particular concern are montane species where recent evidence has identified large-scale shifts in plant communities, with cold adapted species declining and warm-adapted species increasing (Gottfried et al. 2012). Montane and alpine species are likely to be amongst the first impacted by anthropogenic climate change as elevational range shifts extirpate species from

portions of their range (Elsen & Tingley 2015; Kuhn et al. 2016; Lenoir et al. 2008; Guisan & Theurillat 2001). In the UK the montane zone lies above the realized or potential tree line at up to 7-800m in some areas, and as low as 200m in the far North and West of Scotland (Thompson & Brown 1992). A number of species comprising montane scrub habitat, including *Juniperus* and *Salix spp.* have become increasingly range restricted (Mardon 2003; Gilbert & Di Cosmo 2009; Bunce et al. 2014). In particular, dwarf birch (*Betula nana*) is thought to have declined significantly in recent decades as a result of climate change and is classified as nationally scarce in the UK (Groot et al. 1997; Aston 1984), though a robust evaluation has not yet been undertaken. Furthermore, dwarf birch populations are reported to display widely differing performance (e.g. number of catkins, plant size) across extant populations (pers. comm. A. W. Featherstone).

Species and populations however, frequently display idiosyncratic responses to climate change (Parmesan & Yohe 2003; Millar et al. 2004), and environmental conditions in montane regions are highly heterogenous (Engler et al. 2011). Thus it is also possible that climate warming could benefit some montane species or a subset of populations, whilst being detrimental to others (Walck et al. 2011; Pfeifer-Meister et al. 2013; Cleland et al. 2012). In one study, experimental warming of tree birches (*Betula*) extended the growing season and significantly increased the number of male flowers (Nakamura et al. 2016). In another, warming and high nutrient conditions were found to favour arctic communities dominated by dwarf birch, at the expense of other tundra species (Gough et al. 2015). Furthermore, despite broad scale models predicting species extinctions (Randin et al. 2009), fine scale projections find that many species may persist in suitable relict microhabitats or that alpine habitats may buffer plant diversity against warming climate (Scherrer & Körner 2011).

Biotic interactions such as herbivory or pathogens may also influence population fitness (Christie et al. 2014; Olofsson et al. 2009; Fisher et al. 2012). Evidence from a population on the Dundreggan estate, Scotland suggests that dwarf birch inside herbivore exclosures are

significantly larger than those outside and the number of catkins per plant is 13 times greater. Similarly, where plants protrude above the surrounding vegetation, catkin production is 2.5 times greater (Montane scrub survey 2008), despite adaptations to tolerate browsing.

In this study we aim to elucidate the impact of past and projected climate change, and other biotic interactions on the distribution and fitness of *B. nana* in the UK. We achieve this in four stages. First, we collate historical records of *B. nana* to assess evidence of recent decline. Assessing range decline through comparison of historical and contemporary data is often problematic (Tingley & Beissinger 2009; Shaffer et al. 1998), as it is necessary to minimize or control for sampling and observation biases; here we use the closely related species *B. pubescens*, as a control group to correct for uneven sampling effort over recent decades. Second, we compare phenotypic measurements across UK *B. nana* populations, using measurements of populations in Scandinavia as a comparison, to estimate reproductive output. For many montane species in Scotland, there is a lack of evidence of regeneration (Mardon 2003) and populations also appear to recover slowly from disturbance (Bayfield 1979). Thus we test for associations between plant height and grazing pressure, as well as testing for the impact of climate and browsing on catkin production. Third, we project the current and future species distribution of *B. nana* under a range of climate scenarios to estimate habitat suitability, test for uniformity of population response, and guide prioritization of populations for conservation. Finally, we compare SDM derived estimates of habitat suitability across a species' distribution with empirical phenotypic measurements of reproductive output as an indirect proxy for plant fitness. Of the few studies that have evaluated a functional trait together with habitat suitability (summarized in Wittmann et al. 2016), none as far as we are aware have assessed catkin production or germination rate and only one has estimated plant size (Thuiller et al. 2010).

2.2 Methods

2.2.1 Historical records

To assess any trends in the distribution of *B. nana* and to identify a subset of study populations, a search of available literature and records was conducted. A total of 2,245 *Betula nana* records in the UK, ranging from the years 1777 to 2014 were collated from the Botanical Society of Britain and Ireland, Highland Birchwoods consultancy, Trees for Life conservation charity, Scottish Natural Heritage, National Biodiversity Network and the Montane Scrub Action Group databases. To accurately identify changes in *B. nana* distribution, it was necessary to correct for uneven sampling effort. Thus 23,990 records for the same time period were also collated for *B. pubescens*, a closely related species with a distribution that is thought to have remained relatively consistent over the time period considered, whilst also being considerably more abundant and easier to locate. *Betula pubescens* observations were divided into three time periods with equal numbers of records. *B. nana* records were then binned among these periods to a geographical resolution of 0.1 degrees, and plotted using the *ggplot* package (Wickham 2009) in R software (R Development Core Team 2014) and RStudio (RStudio Team, 2015). Populations located during subsequent fieldwork were also added to the final period.

Furthermore, records were obtained for Northern Scandinavia and Ireland. In Scandinavia *B. nana* is extremely abundant and widespread with no detectable change over the time period considered (Karlsson et al. 2007), however only limited observation are available long term with limited resolution, thus the predicted prevalence analysis of Lampinen & Lahti (2016) is reported here (Figure S4). An extensive literature search found no definitive records of living *B. nana* in Ireland, despite several unconfirmed reports. However palenological evidence from pollen cores concludes a high probability of presence since the LGM (Godwin 1957; Birks 1968), and thus infers subsequent extinction.

Based on historical and contemporary records, a range of sampling sites was selected to encompass the extant distribution of *B. nana* in Scotland and Northern England. Where records were old or had low resolution (e.g. 1km grid square) a systematic search method was employed focusing on suitable microhabitats. All UK populations were visited in the summers of 2012, 2013 or 2014, and plants were identified based on leaf morphology and growth structure. In several cases, populations could not be found despite extensive searching. This may be attributed to inaccuracies in the data, cryptic presence at low density or recent population extinction. A sample of Scandinavian populations was selected to encompass a similar geographic scale and sampled in July-August 2013.

2.2.2 Population sampling

A total of 1220 *B. nana* individuals were collected from 29 populations across the species' extant range in the UK and 10 populations from a central portion of the species' European distribution in Scandinavia (Table 2.2). Systematic sampling methods were trialed, however, as *B. nana* individuals were found to occur with clumped distributions, approximately 30-35 plants were haphazardly sampled at each population. For each individual, the following abiotic and phenotypic measurements were recorded and a tissue sample retained for DNA extraction (discussed in subsequent chapters). All measurements were collected by a single investigator (JSB) and up to one other supervised field assistant to ensure consistency.

- Latitude and longitude
- Elevation (Garmin GPS)
- Plant aspect (compass based)
- Plant slope (hand held clinometer)
- Number of male and female catkins.
- Plant height
- Browsing pressure (% of browsed stems).
- Plant area (length of the longest horizontal growing axis X perpendicular width).

- Diameter of the largest stem at ground level.

Due to difficulties in *Betula* DNA extraction from leaves, tissue samples consisted of small twigs 10-20cm long. Attempts were made to minimise resampling of clonal or cryptic individuals by maintaining where possible a minimum distance of 5m between samples. Following guidelines agreed with partner organizations, care was also taken that the material removed amounted to less than 5% of the living plant to avoid damage. Shortly after fieldwork, a section of cambium tissue from each sample was removed, dehydrated and stored on silica gel for DNA extraction. The remaining plant material from each individual was dried in a plant press for 4-8 weeks and then stored on acid-free paper for future reference.

In the years 2013 and 2014, seeds were collected from a subset of 18 populations (9 per year) in Scotland to assess germination rates. Catkins were placed in labeled glassine envelopes and dried for 3-5 days before being stored at 4°C for planting the following spring. It should be noted that collected catkins would have been from the previous year's growth, so are not necessarily correlated to the female catkin count also reported in this study. To assay germination, seeds were counted and spread on filter paper in individually labeled petri dishes. Where a large amount of seed was available for a given individual, several petri dishes were used to avoid overcrowding. A thin layer of vermiculite was added to prevent desiccation. For populations assayed in 2014, successfully germinated seedlings were transferred to a nutrient poor soil (similar to their preferred habitat) to assess the number of seedlings that survived to 100 days post germination.

2.2.3 Species Distribution Modeling

To determine the environmental variables influencing the current and future distribution of *B. nana* we developed an ecological niche model based on the following approach. 763 high-resolution (<100m) *B. nana* records from the period 1960-present were retained, and resampled to a grid size of 0.0083 degrees (approx. 30 seconds of arc, equivalent to a grid

~900 x ~500m) with duplicates removed. The analysis was limited to records from this time period to roughly correspond to the period for which ‘present time’ climatic models were built.

Nineteen bioclimatic layers were obtained from the WorldClim database [www.worldclim.org] at 30 arc-second resolution (Hijmans et al. 2005). This includes 11 temperature and eight precipitation derived variables that are considered more biologically meaningful reflecting annual trends, seasonality and limiting environmental factors. In addition, we utilized WorldClim high resolution elevation data to compute slope and aspect terrain characteristics based on the eight neighboring cells using the *Raster* package (Hijmans & Etten, 2012) in R software. These variables are good indicators of soil moisture, erosion, wind and solar radiation (Hoersch et al. 2002). Elevation was excluded as it is highly correlated with temperature variables (Parolo et al. 2008) which respond to climate change whilst elevation does not, making the former more useful under future climate projections. To avoid over fitting we removed variables where the data was highly correlated (collinearity >0.7), retaining 10 of the most biologically significant variables for analysis (Table 2.1).

Ecological niche model (ENM) generation was conducted using MaxEnt (Phillips et al. 2006) within the *dismo* package (Hijmans et al. 2011). MaxEnt is a commonly used ENM algorithm for presence-only data that has been shown to perform well in comparison with other approaches (González-Irusta, J. et al. 2015; Tarkesh & Gottfried 2012; Elith et al. 2011). Whilst MaxEnt models alone are inappropriate to predict probability of occurrence, they do provides a useful measure of habitat suitability (Royle et al. 2012). ENM derived habitat suitability index (HSI) has also been shown to be highly correlated with species abundance across a range of scales and models (Weber et al. 2016).

Table 2.1 Bioclimatic and abiotic variables included in this study.

Bioclimatic Variables	Retained
BIO1 = Annual Mean Temperature	X
BIO2 = Mean Diurnal Range	X
BIO3 = Isothermality	X
BIO4 = Temperature Seasonality	X
BIO5 = Max Temperature of Warmest Month	X
BIO6 = Min Temperature of Coldest Month	
BIO7 = Temperature Annual Range	
BIO8 = Mean Temperature of Wettest Quarter	X
BIO9 = Mean Temperature of Driest Quarter	X
BIO10 = Mean Temperature of Warmest Quarter	
BIO11 = Mean Temperature of Coldest Quarter	
BIO12 = Annual Precipitation	X
BIO13 = Precipitation of Wettest Month	
BIO14 = Precipitation of Driest Month	
BIO15 = Precipitation Seasonality	
BIO16 = Precipitation of Wettest Quarter	
BIO17 = Precipitation of Driest Quarter	
BIO18 = Precipitation of Warmest Quarter	
BIO19 = Precipitation of Coldest Quarter	
Aspect (derived from elevation)	X
Slope (derived from elevation)	X

We performed analyses with 50 subsampled replicates, jackknife tests of variable importance and a random test percentage of 25% retained for evaluation. Model performance was assessed using the area under the receiver operating characteristic curve (AUC), which ranges from 0.5, no better than random, to 1, perfect discrimination, and examination of variable response curves (Figures S1, S2). In addition, binomial tests of omission were used to confirm whether models differed from a random distribution. A species occurrence threshold to assess changes in species occupied area was defined by 'maximum training sensitivity plus specificity' (Liu et al. 2016), which returned broadly intermediate values in comparison with other threshold metrics. We also performed model estimation for *B. pubescens* and used the method of Warren et al. (2008) to assess niche overlap with the statistic *I*. This approach sums pairwise difference between two predicted distributions, and is potentially more accurate than

simple geographical range overlap as it does not require an arbitrary presence threshold. Finally, in an assessment of MaxEnt's performance in ranking environmental variables by their importance in species range delimitation, Searcy & Shaffer (2016) found that the ranking captures meaningful biologically important factors successfully; thus we report variable ranking for both species.

2.2.4 Projecting future species distributions

Eight further datasets consisting of the same selected variables were then generated under four different emission scenarios defined by the Intergovernmental Panel on Climate Change (IPCC 2014b) at each of two future time periods (averaged for the years 2045-2065 and 2081-2100) to permit future estimation of *B. nana* distribution (Table 2.2). These projections were derived from the Community Climate System Model (CCSM4) (Gent et al. 2011). To assess range overlap and inform our analysis of introgression (Chapter 4), we modeled projected future distribution of *B. pubescens*, using an identical approach with 3087 high-resolution observations from the same time period.

In addition, we characterized the species' niche centroid following the approach of Manthey et al. (2015), by computing a principal component analysis of our retained environmental data layers. The first five axes explained 95.8% of the variation, thus the mean score along these axis was treated as the niche centroid. This enabled us to define the Euclidian distance between each population and the niche centroid. We tested for a relationship between distance to niche centroid and phenotypic measurements in a similar way to habitat suitability.

2.2.5 Comparison of UK and Scandinavian populations

To evaluate broad phenotypic differences between study regions, population-based boxplots of plant height, plant area and stem diameter were plotted and *t* tests used to compare means between the UK and Scandinavia. We also compared total catkin counts between study regions, counts of male and female catkins separately, and counts corrected for plant area

(based on the assumption that populations with larger mean plant size may produce more catkins). Linear regression was performed to evaluate the influence of browsing pressure on mean population plant height. Similarly, mean population browsing percentage was plotted against latitude to assess evidence of a geographical pattern.

Table 2.2 Projected change in global mean surface air temperature, relative to the period 1986–2005. Adapted from (IPCC 2014a).

Scenario	2046-2065	2081-2100
	Mean °C (Likely range)	Mean °C (Likely range)
RCP2.6	1.0 (0.4 to 1.6)	1.0 (0.3 to 1.7)
RCP4.5	1.4 (0.9 to 2.0)	1.8 (1.1 to 2.6)
RCP6.0	1.3 (0.8 to 1.8)	2.2 (1.4 to 3.1)
RCP8.5	2.0 (1.4 to 2.6)	3.7 (2.6 to 4.8)

2.2.6 Catkin production and habitat suitability index

To assess the potential for MaxEnt models to predict plant performance in wild populations, we used reproductive output as a proxy for plant fitness. Reproductive output metrics including catkin or fruit production, seeds per catkin, germination rate and seedling growth have been used extensively to assess fitness in crosses and common garden experiments (Schweitzer et al. 2002; Fritz et al. 2006; Gramlich & Horandl 2016). Thus we fitted a generalized linear model with a quasipoisson error distribution to test for a relationship between present time HSI estimates and mean population catkin counts. We also tested for a relationship between HSI and mean germination rates using a negative binomial error distribution. Here we are explicitly testing the hypothesis that plants displayed greater reproductive output in locations with a higher habitat suitability index.

2.3 Results

2.3.1 Historical records

Database searches revealed 2,245 *B. nana* records for the UK, ranging from years 1777 to 2012. For the same time period, 23,990 *B. pubescens* records were available. Normalization against *B. pubescens* records resulted in three periods of equal sampling effort; 1777-1988, 1989-1999 and 2000-2012 (Figure 2.1). Distribution maps show a decline and contraction of *B. nana* range to three main clusters in the Cairngorms, Rannoch and Ross & Cromarty regions. A small number of relict populations still persist in Northern England.

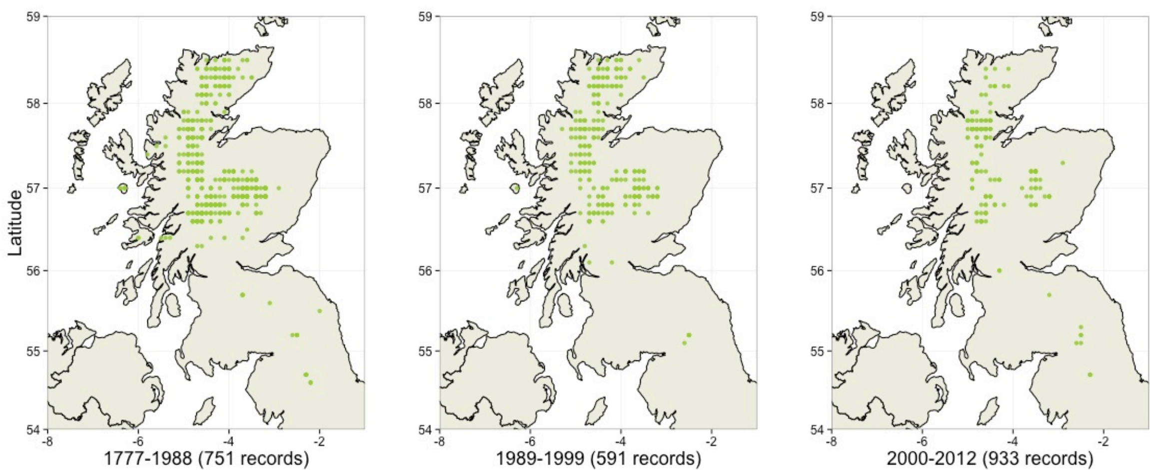


Figure 2.1 Historical records of dwarf birch normalized to three periods of equal sampling effort. Distribution shows progressive contraction of range to the Scottish highlands.

2.3.2 Population sampling

The location of *B. nana* populations sampled in the UK and Scandinavia are shown are reported in Figure 2.2. In UK populations, all individuals found of sufficient size were sampled up to a total of 35 individuals per population (except for populations LG and PC in which all individuals were exhaustively sampled to address another research question). In Scandinavia, where *B. nana* is widespread and occurs in large continuous populations, ~32 samples were collected at each sample site. The remaining populations in the UK reflect a pattern of increased fragmentation and decreasing size to the south of the *B. nana* range. This southern

decline in *B. nana* populations was qualitatively further evidenced by the fact that several populations recorded by other <40 years ago could not be located, suggesting possible local extinction.

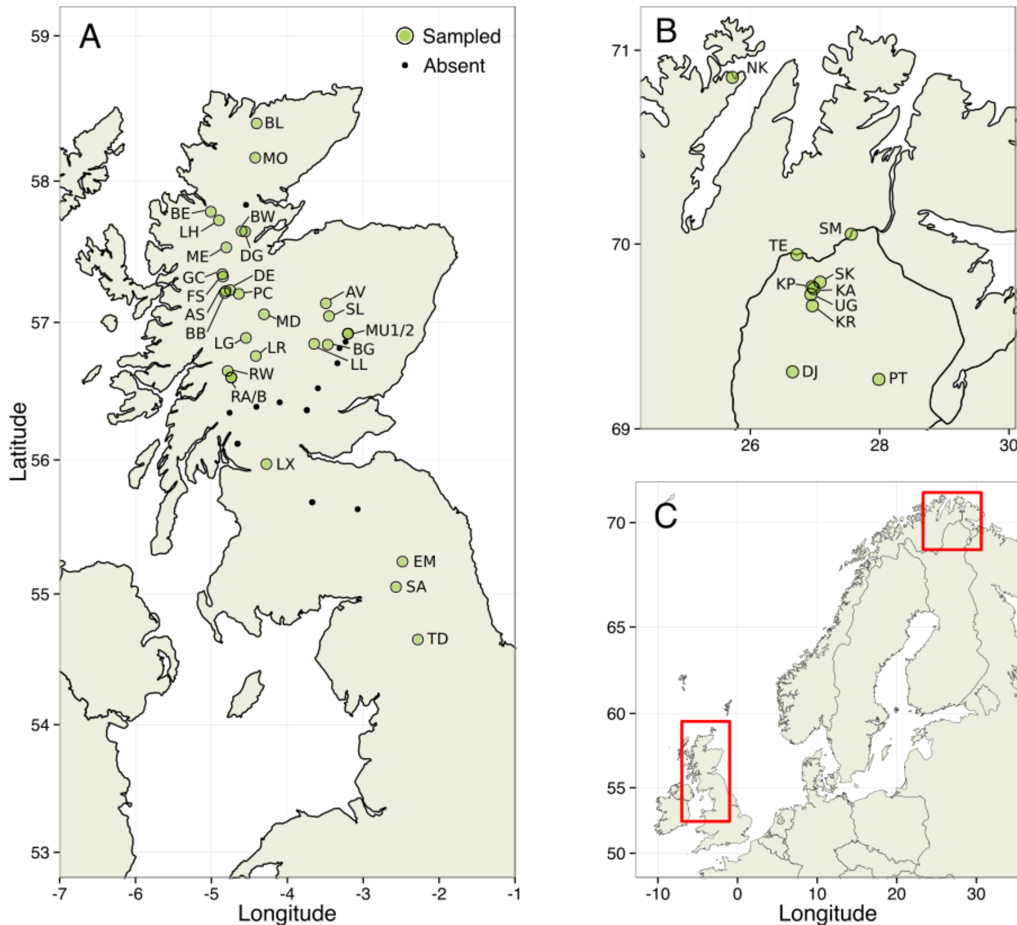


Figure 2.2 A) Sampling locations in England and Scotland (UK). B) Sampling locations in Finland and Norway (Scandinavia). C) Map of Northern Europe identifying study regions. Green circles are sampled populations. Black points denote populations from historical records that could not be relocated, and thus may be locally extinct.

2.3.3 Species Distribution Modeling

After filtering to reduce spatial sampling bias, we retained high-resolution records from 78 grid squares. The *B. nana* MaxEnt model performed well at predicting habitat suitability with high mean test AUC (0.946 ± 0.008) and a low mean test omission rate (0.09, p <0.001) at a logistic threshold of occurrence of 0.193. Five variables contributed substantially to the model (>5%);

these were annual mean temperature (34.9%), maximum temperature of the warmest month (22.1%), mean diurnal range (14.8%), isothermality (14.6%) and annual precipitation (7.3%). Examination of model variable response curves identified how the logistic prediction changes as each environmental variable is varied, keeping other environmental variables at their average sample value. Annual mean temperature was characterised by a severe decline in probability of presence as temperature rose above 7°C. Similarly Isothermality also declined from 3.3 - 4.2°C, displaying an approximately negative exponential distribution. Maximum temperature of the warmest month varied between 13 - 16.5°C, but declined significantly thereafter. Annual precipitation reported the highest probability of presence at approximately 900mm, declining steadily with increasing precipitation over a broad range of values, with a severe decline beyond 1900mm. Mean diurnal range was unusual in that it displayed a positively skewed normal distribution with an optimum value of approximately 6.4°C, the model appeared more tolerant to values slightly higher than this, and less to values lower. Variables with very small contributions are harder to interpret, but do not substantially influence the results of the model. Evaluation of permutation importance identified the same top five variables, but ranked them differently (Figure 2.3, Table 2.3).

Values for the *B. pubescens* model were lower (AUC 0.784; omission 0.24, $p < 0.001$; threshold 0.42), but nevertheless showed high discriminatory capacity. Lower AUC is also expected for a more widely ranging species (Lobo et al. 2008), which is consistent with our ecological knowledge of *B. pubescens* (Godwin 1957). The highest contributing variable was temperature seasonality, which increased across the range of values, and was considerably more important in the *B. pubescens* model than for *B. nana*. Maximum temperature of the warmest month displayed a negatively skewed distribution, with the peak value approximately 19°C, substantially higher than the optimum for *B. nana* as expected for a montane versus upland species. Mean diurnal range appeared to show a bimodal distribution which may be due to poorly characterized interactions with other variables (Elith et al. 2010). Isothermality declined

rapidly as values exceeded 3.8°C, and conversely annual mean temperature displayed a positive exponential curve up to a peak value of 8.5°C, before declining at an apparent biological threshold. Thus the overall pattern of variable importance differed between the two species (Figure 2.3, Table 2.3), variable response curves for *B. pubescens* are reported in Figure S2.

Whilst the current predicted habitat suitability maps for *B. nana* is highly concordant with field observations, future projections show significant declines across the species' range with limited persistence in higher altitude areas (Figure 2.3). By contrast the current predicted habitat suitability maps for *B. pubescens* show a much broader distribution, but with similar trends (Figure 2.3). In the future, overall decline in *B. pubescens* across the study area is less severe with localized reduction in habitat suitability particularly in lowland areas.

Habitat suitability for all species records under the four future scenarios are plotted in Figure 2.4. This identifies an overall negative trend in habitat suitability for *B. nana*, whilst also showing an overall consistent or positive trend for *B. pubescens*. Our models suggests that under the highest emission scenario (rcp85, 2070), suitable *B. nana* habitat may be reduced to ~1% of the current extent. The total change in area, range overlap and niche overlap is given in Table 2.4.

Figure 2.3 MaxEnt models of *Betula nana* and *B. pubescens* under current and future climate scenarios. Black points indicate species distribution records. Red points indicate sampled locations included in this study.

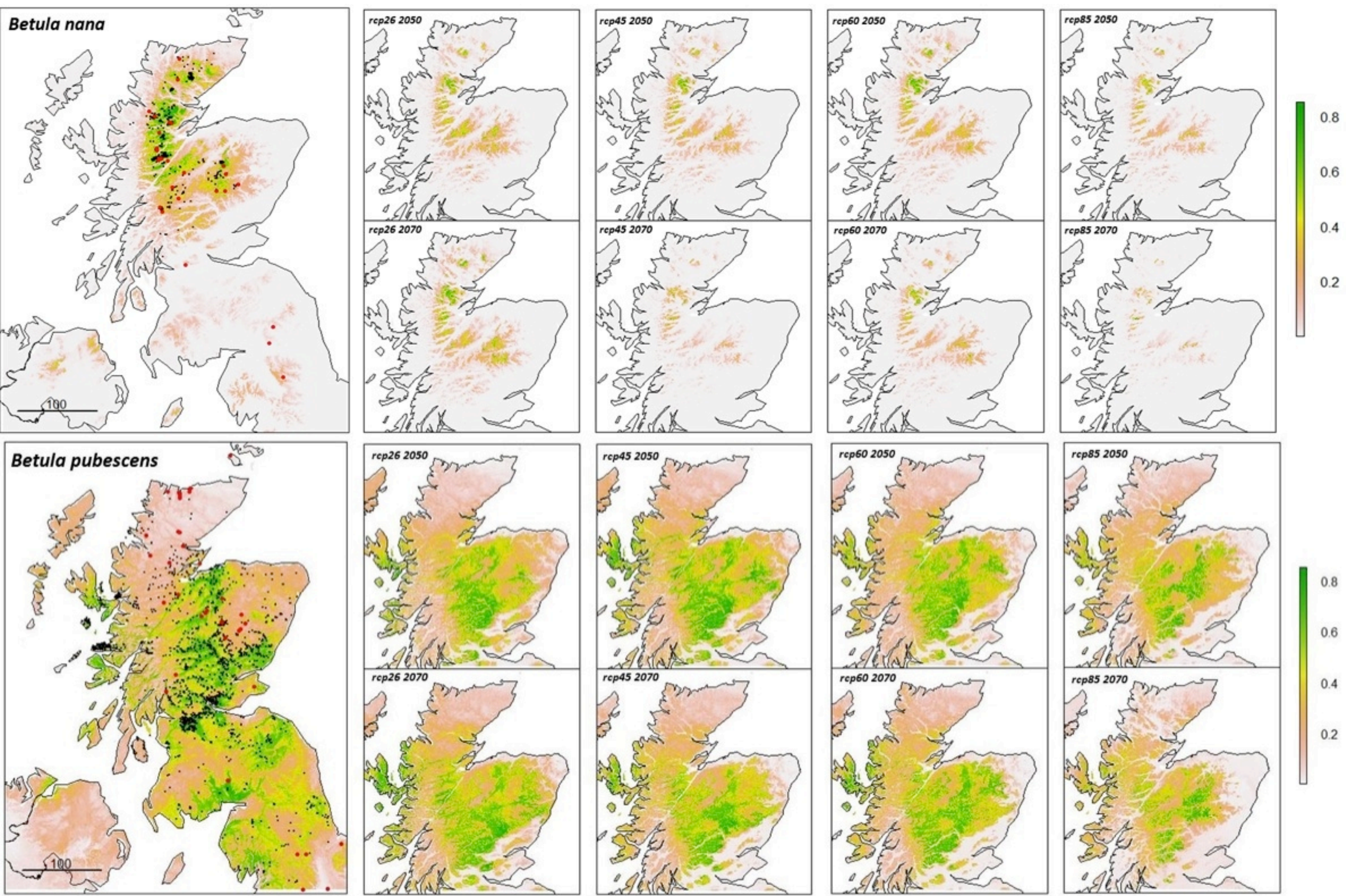


Table 2.3 Environmental variable importance for estimating MaxEnt species distribution models. Associated variable response curves are plotted in Figures S1 and S2.

Variable	Percent	Jackknife
Betula nana		
Annual Mean Temperature	34.9	5.6
Max Temperature of Warmest Month	22.1	32.2
Mean Diurnal Range	14.8	17.1
Isothermality	14.6	26.5
Annual Precipitation	7.3	8
Slope	2.8	3.4
Mean Temperature of Driest Quarter	1.6	5
Temperature Seasonality	1.4	1.3
Mean Temperature of Wettest Quarter	0.3	0.8
Aspect	0.2	0.2
Betula pubescens		
Temperature Seasonality	25.5	27.1
Mean Diurnal Range	21	30.4
Max Temperature of Warmest Month	20.8	5.1
Mean Temperature of Driest Quarter	10.2	5
Isothermality	8.4	8.9
Annual Mean Temperature	5.9	10.8
Annual Precipitation	3.5	6.6
Mean Temperature of Wettest Quarter	2.2	3
Slope	2.1	2.3
Aspect	0.5	0.6

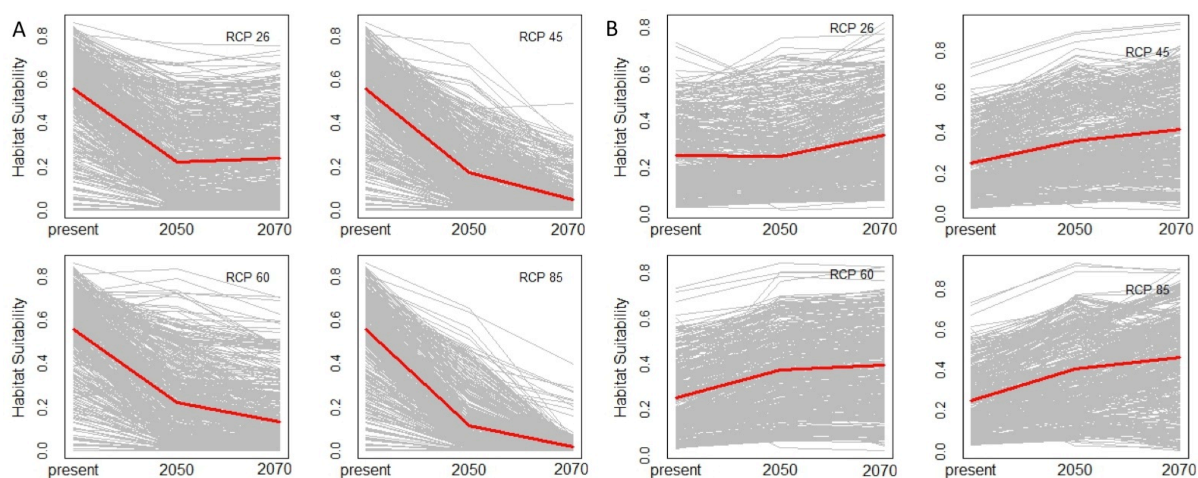


Figure 2.4 Changes in habitat suitability under four future climate scenarios for *Betula nana* and *B. pubescens.* The red line indicates overall mean for all recorded locations.

Table 2.4 Change in habitat area for *B. nana* and *B. pubescens* under future climate scenarios.

Species	Current area(Km^2)	2046-2065				2081-2100			
		RCP26	RCP45	RCP60	RCP85	RCP26	RCP45	RCP60	RCP85
B. nana	11415	2799	1774	2783	952	3021	463	1406	128
B. pubescens	28824	14647	21523	28760	19335	25491	13717	20857	12200
Range overlap	2061	553	496	878	270	763	101	459	43
Niche overlap (I)	0.54	0.51	0.53	0.54	0.56	0.52	0.6	0.56	0.66

2.3.4 Phenotypic comparison of UK and Scandinavian populations

Sampled populations and summarized geographic, abiotic and phenotypic measurements are reported in Table 2.5. Overall, abiotic measurements identified concordance in aspect and slope specificity for *B. nana* in both the UK and Scandinavia (Figure 2.5), with the majority of populations occurring on flat to moderately sloping ground with a northerly aspect. In many cases, even where the landscape was flat, clumps of *B. nana* tended to be localized to the northern side of small terrain features such as hummocks or rocks.

Plant height, area and stem diameter are plotted in Figure 2.6 with samples ordered by region and decreasing latitude. Mean plant height per population was found to be significantly higher in Scandinavia ($t = -4.44$, $df = 11.6$, p-value = <0.001), though in both regions there is high variability most likely due to grazing pressure and exposure in various microhabitats. For example, the very low mean height for the Scandinavian Nordkapp population is probably due to severe exposure at this locality. The considerable height of some of the Southernmost relict UK populations also noteworthy. Plant area appeared to display a similar pattern with Scandinavian individuals typically being larger; however no significant difference was detected between regions ($t = -2.36$, $df = 26.8$, $P = 0.19$). Overall however, the number of very large plants is much reduced in the UK except at the Southern range edge. Finally, mean stem diameter was significantly greater in Scandinavia ($t = -5.18$, $df = 25.34$, $P < 0.001$), though several notably large individuals were found at Loch Muick and River Avon, UK.

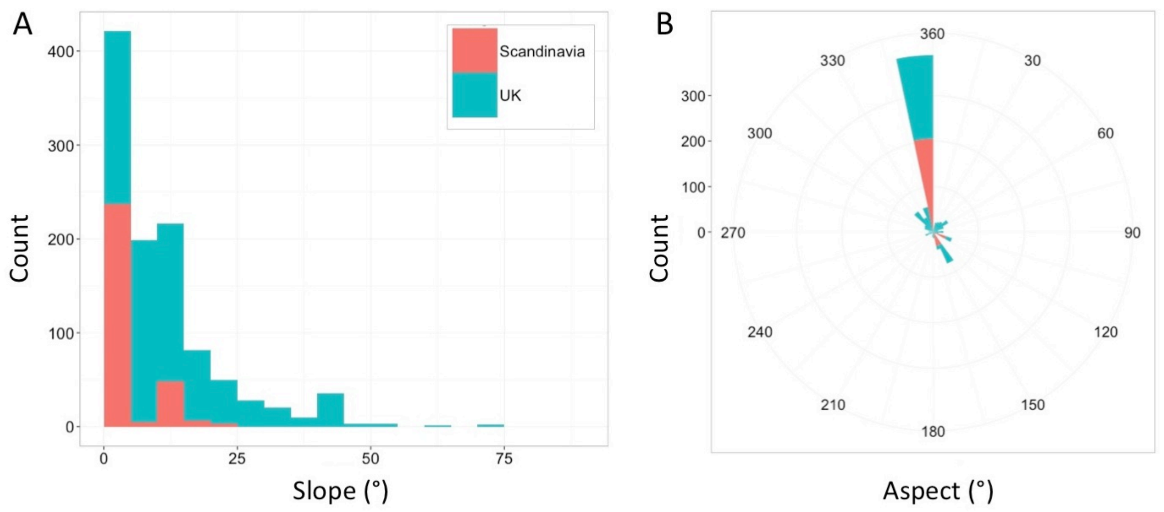


Figure 2.5 Occurrence of *B. nana* across a range of habitat slope and aspect values in the UK and Scandinavia.

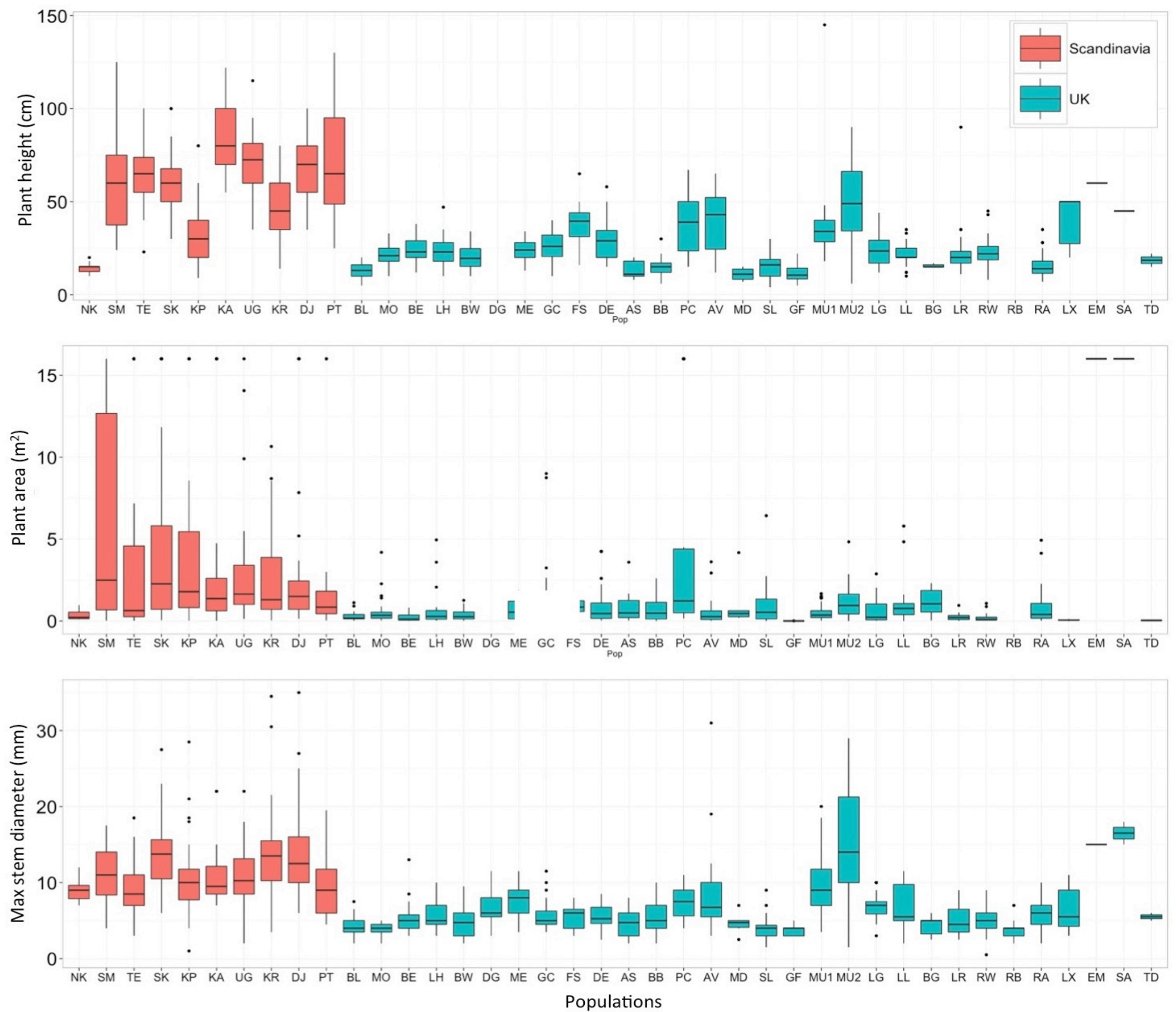


Figure 2.6 Boxplot of phenotypic traits for *B. nana* across Scandinavia and the UK. Populations are ordered by latitude with more northerly populations on the left hand side.

Table 2.6 Samplin **Table 2.5** Sampling locations and phenotypic data summary for sampled *B. nana* populations.

Population	Pop ID	Latitude	Longitude	Sampled[1]	Altitude (m a.s.l.)	Male catkins	Fem. catkins	Total area (m^2)	Height (cm)	Relative Height (cm)	Browsing (%)	Stem diameter (mm)
UK												
Ben Loyal	BL	58.401	-4.404	31	300	0.17	2.43	0.27	13.13	-3.57	41	4.25
Meall Odhar	MO	58.163	-4.423	32	404	1.03	28.34	0.61	21.69	-0.54	16	3.88
Beinn Enaiglair	BE	57.786	-5.009	27	480	1.52	15.33	0.23	24	4.26	31	5.11
Luichart	LH	57.725	-4.9	33	268	1.42	12.39	0.62	23.52	9.24	40	5.77
Ben Wyvis	BW	57.65	-4.602	34	482	3.13	16.57	0.35	19.77	2.86	19	4.8
DJG Ben Wyvis	DG	57.646	-4.556	21	472	-	-	-	-	-	-	6.86
Loch Meig	ME	57.534	-4.804	26	450	5.42	9.96	0.7	23.67	3.33	36	7.67
Glen Cannich	GC	57.345	-4.856	33	455	3.84	37.65	1.33	26.26	6.61	9	5.69
Faskanyle	FS	57.327	-4.848	31	486	32.31	57.19	0.89	38.1	12.27	12	5.54
Dundreggan E.	DE	57.231	-4.754	30	448	24.83	56.31	0.92	28.97	12.93	0	5.5
An Suidhe	AS	57.224	-4.812	31	661	0.71	0.88	0.89	12.76	-2.65	6	4.5
Beinn Bhreac	BB	57.211	-4.823	33	500	5.15	4.88	0.72	15.15	-1.41	40	5.2
Portclair	PC	57.204	-4.639	44	478	8.16	61.63	8.71	36.58	17.37	20	7.3
River Avon	AV	57.137	-3.491	31	549	9	12.75	0.58	38.75	6.07	29	8.48
Monadhliaths	MD	57.058	-4.307	36	712	0	0.33	1.03	11	-5	25	4.67
Meall an tslugain	SL	57.045	-3.451	31	633	1.42	0.23	0.93	15.08	-4.83	51	3.88
Loch Muick 1	MU1	56.92	-3.198	31	492	1.45	0.94	0.54	37.52	-1.13	40	9.61
Loch Muick 2	MU2	56.918	-3.205	32	517	0.69	1.19	1.25	50.06	14.88	41	14.59
Loch Laggan	LG	56.889	-4.545	52	364	0.77	6.77	0.64	23.73	7.5	43	6.81
Loch Loch	LL	56.846	-3.647	32	673	11.53	6.84	0.99	21.72	0	43	6.75
Ben Gullabin	BG	56.84	-3.467	7	594	0.14	0	1.18	15.57	-0.71	66	4.29
Loch Rannoch	LR	56.758	-4.415	30	499	8.71	25.46	0.25	23.04	6.43	14	5.13
Rannoch West	RW	56.65	-4.785	32	306	3.75	3.28	0.19	22.72	6.77	38	5.08

Table 2.5 Cont.

Population	Pop ID	Latitude	Longitude	Sampled[1]	Altitude (m a.s.l.)	Male catkins	Fem. catkins	Total area (m^2)	Height (cm)	Relative Height (cm)	Browsing (%)	Stem diameter (mm)
Rannoch Moor B	RB	56.603	-4.74	32	304	0	2.1	-	-	-	13	3.89
Rannoch Moor A	RA	56.603	-4.738	30	295	3.93	12.56	0.88	15.7	0.96	13	5.76
Lennox	LX	55.97	-4.276	10	164	2	5.88	-	41	15	-	6.5
Emblehope	EM	55.244	-2.483	6	448	50	300	25	60	40	10	15
Spadeadam	SA	55.053	-2.568	6	275	0	0	-	45	25	-	15
Teesdale	TD	54.654	-2.28	2	499	0	4	-	18.5	12.5	18	5.5
Scandinavia												
Nordkapp	NK	70.864	25.721	30	25	0.2	0.8	-	14.4	-	-	9
Sirbma	SM	70.052	27.562	32	79	2.2	21.8	5.48	59.3	50.2	14.7	11.1
Tenontie	TE	69.944	26.723	30	96	0.7	4.4	3.43	63.9	51.8	11	9.1
Skalluvaara	SK	69.799	27.079	32	222	3.7	23.3	4.28	59.8	49.3	19.5	13.9
Kevo_Plateau	KP	69.774	26.956	32	310	0.1	2.4	4.31	33.4	27.3	26.9	10.7
Kevojarvi	KA	69.763	26.981	32	128	1.2	1.2	3.27	83.9	58.3	8	10.7
Gearddosjavri	UG	69.732	26.937	24	119	1.3	12	3.23	71.7	53.8	8.5	11.2
Kevo_Reserve	KR	69.672	26.961	31	225	1.2	16.1	2.68	47.2	39.6	20	14.1
Kotilampi	DJ	69.313	26.654	29	217	1.6	11	3.01	68.3	57.6	21.1	14.3
Partakko	PT	69.273	27.988	28	127	0.4	4.9	2.41	70.6	57.5	11.2	9.4

[1] A small number of samples were excluded from phenotypic analysis due to missing data, resulting from inclement weather conditions curtailing fieldwork.

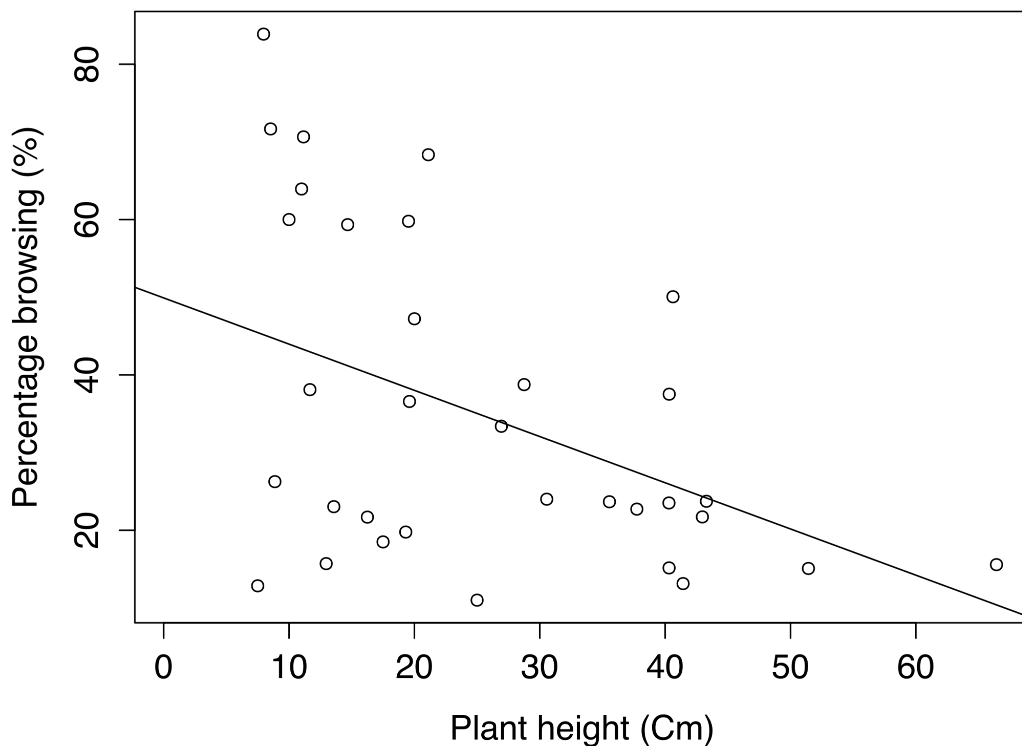


Figure 2.7 Scatter plot of mean population plant height against mean percentage browsing for *B. nana* individuals across all populations.

Percentage of shoots browsed varied substantially across populations in the UK, but was significantly lower in Scandinavia (t = 3.030, df = 32.047, P < 0.005). Browsing pressure also was significantly correlated with plant height ($F_{1,32}$ = 7.8, R^2 = 0.17, P = 0.008) suggesting higher browsing pressure was associated with shorter plants (Figure 2.7). No significant relationship was detected with latitude in the UK. Overall, principal component analysis of population means identified two main groups, corresponding to the two study regions (Figure 2.6). The first two axes together describe 75.18% of the variation, with total area and stem diameter separating Scandinavia from the UK and catkin production separating populations in the UK. Similarly overall variation on both axes was lower in Scandinavia.

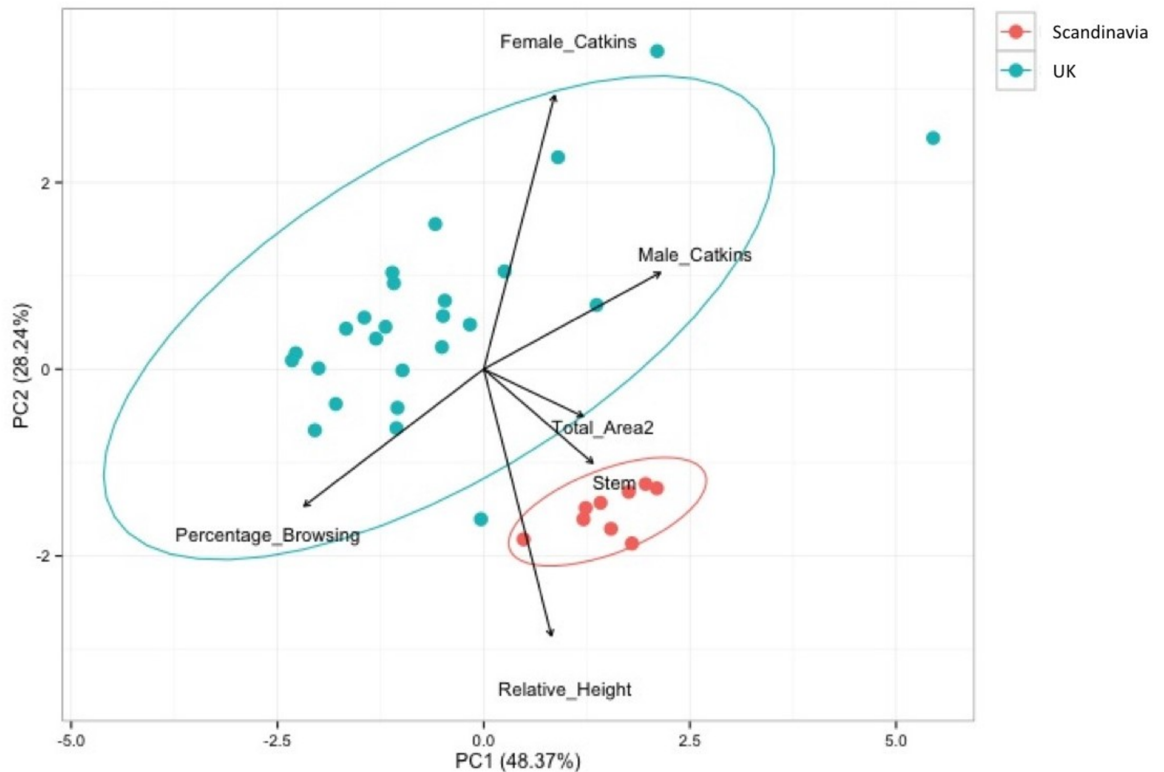


Figure 2.8 Principal component analysis of *B. nana* phenotypic characteristics in the UK and Scandinavia.

2.3.5 Catkin production and germination success

Catkin production varied substantially across individuals and populations, ranging from 0 to >450 per plant. Raw population means for total catkin count did not significantly differ between the UK and Scandinavia (t = 1.44, df = 31.47, P = 0.16). However, after correcting for plant area (which is on average substantially larger in Scandinavia), a highly significant difference emerged (t = 3.21, df = 26.2, P = 0.004). Results were also significant for corrected male and female catkin counts analysed separately (data not shown). Population counts are summarized in Figure 2.9.

Germination success was assayed in 190 individuals from UK populations, with an overall germination rate of 7.6% and the percentage 100-day survivability of 6.12% (Table 2.6). We found substantial variation across populations, with germination rate appearing to be the

more important limiting factor rather than 100-day survivability in this limited sample of populations.

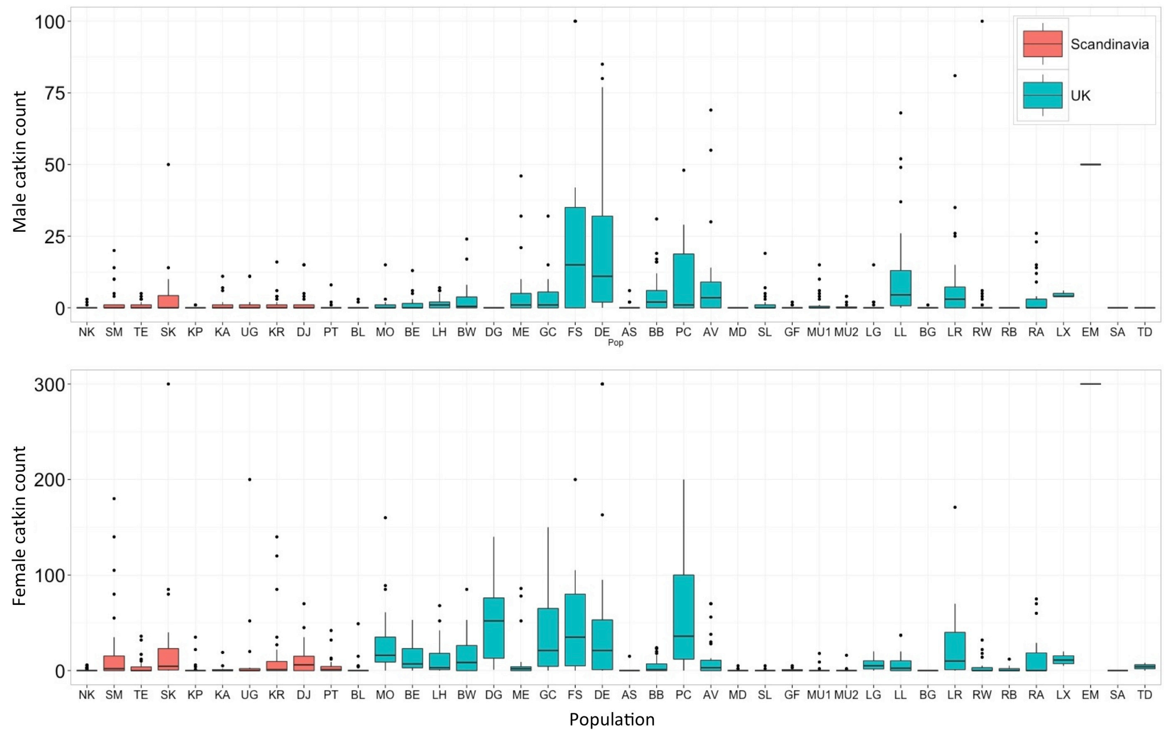


Figure 2.9 Boxplots of male and female catkin counts across all populations

Table 2.6 Summary of seed sampling, germination success and 100-day survivability data for *B. nana* in the UK.

Year	Populatio	Individua	Seeds	Germinate	Germ.	100-Day	Surv. %
2013	AV	13	438	1	0.23	-	-
2013	BB	7	102	2	1.96	-	-
2013	BL	2	35	0	0.00	-	-
2013	DE	17	540	24	4.44	-	-
2013	GC	8	833	68	8.16	-	-
2013	LL	10	187	0	0.00	-	-
2013	LR	6	63	0	0.00	-	-
2013	ME	3	67	0	0.00	-	-
2013	MU	8	151	0	0.00	-	-
Total		74	2416	95	3.93	-	-
2014	DJG	21	1345	134	9.96	27	2.01
2014	FS	23	492	89	18.09	86	17.48
2014	LG	31	310	16	5.16	15	4.84
2014	LX	5	230	0	0.00	0	0.00
2014	RA	2	31	1	3.23	0	0.00
2014	RB	3	21	0	0.00	0	0.00
2014	PC	28	672	101	15.03	77	11.46
2014	TD	2	14	0	0.00	0	0.00
2014	EM	1	250	5	2.00	1	0.40
Total		116	3365	346	10.28	206	6.12

Table 2.7 Niche distance and HSI values for UK *B. nana* populations under current and future conditions.

ID	Latitude	Longitude	Niche D	Current	2050				2070			
					RCP26	RCP45	RCP60	RCP85	RCP26	RCP45	RCP60	RCP85
BL	58.40	-4.40	3.55	0.38	0.06	0.09	0.13	0.12	0.07	0.01	0.04	0.00
MO	58.16	-4.42	2.49	0.45	0.21	0.08	0.23	0.12	0.28	0.02	0.12	0.01
BE	57.79	-5.01	2.51	0.37	0.04	0.02	0.06	0.02	0.11	0.03	0.01	0.01
LH	57.73	-4.90	2.60	0.54	0.02	0.01	0.02	0.02	0.05	0.01	0.01	0.01
BW	57.65	-4.60	1.03	0.77	0.48	0.51	0.60	0.57	0.55	0.07	0.59	0.02
DG	57.65	-4.56	0.68	0.75	0.31	0.25	0.49	0.05	0.32	0.02	0.12	0.01
ME	57.53	-4.80	2.63	0.57	0.22	0.08	0.31	0.02	0.20	0.02	0.06	0.01
GC	57.35	-4.86	2.82	0.51	0.16	0.08	0.24	0.02	0.31	0.01	0.07	0.01
FS	57.33	-4.85	1.90	0.66	0.15	0.21	0.35	0.15	0.35	0.04	0.12	0.01
DE	57.23	-4.75	1.77	0.81	0.31	0.38	0.45	0.08	0.37	0.03	0.07	0.01
AS	57.22	-4.81	2.57	0.77	0.47	0.52	0.49	0.31	0.51	0.21	0.30	0.06
BB	57.21	-4.82	1.71	0.66	0.16	0.21	0.18	0.19	0.20	0.02	0.16	0.01
PC	57.20	-4.64	1.81	0.54	0.38	0.06	0.11	0.03	0.12	0.01	0.04	0.00
AV	57.14	-3.49	2.40	0.59	0.18	0.12	0.09	0.03	0.28	0.02	0.06	0.00
MD	57.06	-4.31	3.83	0.49	0.40	0.29	0.24	0.16	0.30	0.09	0.15	0.03
SL	57.05	-3.45	2.52	0.59	0.16	0.27	0.32	0.10	0.27	0.03	0.21	0.01
MU	56.92	-3.20	2.83	0.17	0.05	0.04	0.12	0.01	0.04	0.00	0.04	0.00
MU	56.92	-3.21	2.94	0.10	0.02	0.00	0.02	0.00	0.01	0.00	0.00	0.00
LG	56.89	-4.55	2.89	0.35	0.02	0.01	0.01	0.00	0.02	0.00	0.01	0.00
LL	56.85	-3.65	2.70	0.57	0.28	0.09	0.19	0.08	0.25	0.02	0.13	0.01
BG	56.84	-3.47	2.91	0.58	0.21	0.07	0.26	0.11	0.30	0.02	0.22	0.01
LR	56.76	-4.42	1.98	0.23	0.02	0.02	0.01	0.01	0.02	0.00	0.01	0.00
RW	56.65	-4.79	3.56	0.61	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RB	56.60	-4.74	3.42	0.51	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RA	56.60	-4.74	3.42	0.51	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LX	55.97	-4.28	4.42	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EM	55.24	-2.48	2.98	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SA	55.05	-2.57	4.43	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TD	54.65	-2.28	2.47	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

2.3.6 Habitat suitability index

Current HSI estimates for UK *B. nana* study populations ranged from 0.0006 (LX) to 0.81 (DE) (Table 2.7) providing a broad range of values with which to assess reproductive output and germination as a proxy for plant fitness. We found a highly significant positive relationship between HSI and total catkin count ($F_{1,26}$=7.50, P=0.011) as well as HIS and percentage germination ($Z_{1,16}$=2.24, P=0.025) (Figure 2.10). The observed relationship for total catkin count was also significant after correction for plant area ($F_{1,26}$=2.74, P=0.025) (Figure S3). Surprisingly, we found a significant negative relationship between HSI and mean population height ($F_{1,25}$=10.48, P=0.003, R^2=0.27) and HSI and mean population stem diameter

($F_{1,27}$=14.37, P<0.001, R^2=0.32). No significant relationship was found for total plant area. Ecological distance to niche centroid was smallest in populations on Ben Wyvis (1.03) mountain, and highest in the Southern range edge populations LX (4.42) and SA (4.43), values are reported in Table 2.8. Increasing distance to niche centroid was also a negative predictor of mean population catkin count ($F_{1,24}$=4.459, P=0.04, R^2=0.12), but only explained a small amount of total variation. There was no clear relationship with height or plant area.

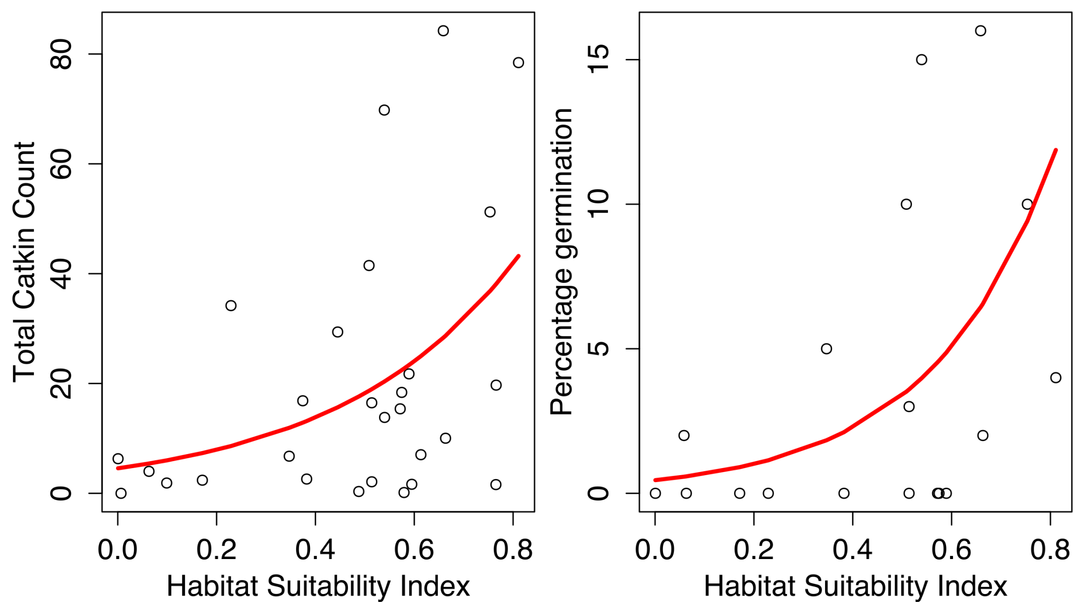


Figure 2.10 GLM of mean population total catkin count against the MaxEnt derived habitat suitability index showing that more suitable sites are associated with a greater number of catkin and higher percentage germination. A comparative plot of catkin count corrected for plant area is given in Figure S3.

2.4 Discussion

The montane tree *B. nana* has shown a clear range decline across the Scottish Highlands and Northern England over recent decades. This is consistent with the historic northward range shift observed since the last glacial maximum described in Wang et al. (2014). The fact that records mainly from lowland areas on the Southern range edge could not be located would also appear to support a pattern of northward shift. However, though no latitudinal pattern in percentage browsing was found in the UK, factors such as agriculture and heterogeneous grazing pressure cannot be ruled out.

For several measured phenotypic traits, *B. nana* populations in Scandinavia and the UK show significant differences. Scandinavian plants were on average significantly taller and with greater mean stem diameter that may indicate a more varied age demography and/or lower herbivore pressure. Observational evidence also clearly shows that *B. nana* is more abundant in Scandinavia (Figures 2.1, S4), although it still persists as locally common in specific parts of the UK (e.g. West of Loch Ness and around Rannoch Moor). However the degree of population variation in phenotypic traits is greater in the UK, as shown by greater variation on PCA axes. This potentially results from exposure to greater variation in landscape and environmental characteristics in the more mountainous Scottish highlands.

Catkin production is higher in UK populations, both before and after correction for plant area. Whilst this may be attributed to greater productivity at lower latitudes and higher temperatures (Körner 2012), it could also represent a greater investment in vegetative growth at higher latitudes, which has been reported previously (Holm 1994; Weis & Hermanutz 1993). This does however fail to explain the decline in catkin production at the southern UK range limit, suggesting an inability to maintain sexual regeneration beyond an unknown environmental threshold.

Browsing was a significant negative predictor of catkin production in the UK. Browsing pressure is overall higher in the UK than in Scandinavia, but highly variable, likely due to different management practices on various estates (Smart et al. 2004). Recent research has implicated high deer densities as a limiting factor in the regeneration of upland birchwoods (Tanentzap et al. 2013), particularly at the tree line where slow growth and short maximum height may prevent trees escaping browser pressure (see also analysis of the impact of browsing on montane scrub, Chapter 5).

Four environmental variables including annual mean temperature and maximum temperature of the warmest month contributed significantly to the *B. nana* SDM. Whilst Maximum temperature of the warmest month was important for both *B. nana* and *B. pubescens*, overall, the important variables and variables responses different markedly between the two species models. Despite this, we find that range overlap and niche similarity is likely to increase between the two species. Whilst these models are parameterized on a course resolution, it is however likely that *B. nana* and *B. pubescens* may be able to persist together in the same grid square by occupying different microhabitats. In the future, species distribution models project that UK range decline is likely to continue and become increasingly severe, with almost total range loss possible by the late 21st century under worst-case scenarios. By contrast, *B. pubescens* distribution is projected to shift North-west, but remain of a similar extent in best-case scenarios, and be reduced by up to 50% in the worse case. Theoretical niche overlap remains approximately the same, but realized range overlap (as a proportion *of B. nana* range) increases incrementally with warming.

The overall decline in habitat suitability (see Fig. 2.8A) suggests deterioration of *B. nana* populations. However a longstanding weakness of SDMs is that there is no clear relationship between model-derived values of habitat suitability and organism fitness (Guisan & Thuiller 2005; Stanton-Geddes et al. 2012). Here we have shown that reproductive output (catkin production and seed germination rates) is higher at locations that the SDM identifies as more

favorable. We consider this a substantial empirical validation of model habitat suitability estimates, and it may also indicate biological process related to environmental conditions. For example, Alsos et al. (2003) found apparent deviation from thermal requirements in Svalbard drastically reduced *B. nana* germination rates. Thus, based on these findings we can predict with greater confidence that as habitat suitability declines, extant populations are likely to experience reduced fitness, and accordingly be subject to greater selection pressure. The correlation between fitness and habitat suitability remains untested in all but a very limited number of species (see summary in Wittmann et al. 2016). Consideration of the wider implications suggests that a similar response may be observed in other montane plants.

Finally, the heatmap of habitat suitability under current and future conditions identifies three important points. Firstly, there is a latitudinal trend in the decline of habitat suitability, as would be expected under a warming climate, with southern populations showing the greatest and earliest deterioration. Second, this approach provides a method with which to identify potential refugial populations, for example here a mid-latitude region to the west of Loch Ness (ME,GC,FS DE and AS) as well as Ben Wyvis (BW) show moderate resilience. Thirdly, the response is non-linear particularly under *rcp26*, as some estimates are lower for the 2050 scenarios before recovering slightly for the 2070 period (for example BG – *RCP2.6*).

Overall we conclude an inability for *B. nana* to track changing climate, which has been exacerbated by anthropogenic drivers such as increased grazing. In the near term, management interventions that focus on reducing browsing may have the most positive impact on plant size and reproductive output. In the future, evidence suggests that the disequilibrium between realized range and climate is likely to deteriorate. Nevertheless, habitat suitability declines to different extents across the species distribution, thus prioritizing populations with high favourable future habitat suitability and high catkin production and germination rates may be a prudent management strategy.

Chapter 3: Combining markers with different mutation rates to infer the population genetic consequences of fragmentation in dwarf birch

Summary

Genetic analyses of fragmented tree populations seldom find the loss of genetic diversity predicted by simple population genetic models. Here, we investigate this paradox in populations of dwarf birch (*Betula nana*), which has declined substantially over recent decades and now persists in a handful of relict sites across the UK. We compare these populations with large unfragmented populations in Scandinavia, using genetic markers with differing mutation rates. We combine PCR-generated microsatellites, RADseq generated transition and transversion SNPs, and novel sequence derived microsatellites generated from RADseq reads (RAD-SSRs). The markers all show that the genetic diversity of dwarf birch in the UK is best explained by recent population bottlenecks coinciding with historic shifts in land use, rather than a more ancient bottleneck during the of the UK after the last glaciation. We find that small microsatellite datasets contain as much information as thousands of genome wide SNPs for many applications. We found that combining information from multiple independent marker sets improves confidence in our conclusions; firstly because they vary by many orders of magnitude in mutation rate and therefore are affected differently by genetic drift and gene flow, and secondly because they have different ascertainment biases. Finally, we describe how RAD-SSRs may overcome many of the limitations of PCR-generated microsatellite markers, providing a method to generate a high mutation-rate marker set with a well-specified ascertainment process from existing RADseq data.

3.1 Introduction

Fragmented tree populations often fail to show the diversity loss within populations and differentiation between them that are predicted by population genetics models (Young et al. 1996; Piotti 2009; Bacles & Jump 2011). In numerous studies (Jump & Peñuelas 2006; Finger et al. 2012; Dubreuil et al. 2010; Davies et al. 2013; Ismail et al. 2012) the effects of fragmentation on tree population genetics remain elusive. This discrepancy is termed the forest fragmentation paradox by Kramer et al. (2008). It is important to resolve this paradox since the predicted effects on the populations' genetic composition could impair the conservation of tree genetic diversity (Koskela et al. 2013).

In the recent past, human activities have extensively fragmented the formerly continuous ranges of many woody plants (Kaplan et al. 2009), reducing population sizes, increasing isolation and potentially altering landscape patterns of genetic structure (Fahrig 2003; Piotti 2009; Crispo et al. 2011). Over successive generations, the reduction of local effective population size in fragmented populations would be expected to result in elevated inbreeding and decreased heterozygosity (Ellstrand & Elam 1993). These processes might be expected to culminate in reduced fitness (Buza et al. 2000; Aguilar et al. 2008) with implications for long term viability being most severe where effective population is reduced the most (Ellstrand & Elam 1993; Lacy 1987).

Several solutions to the forest fragmentation paradox have been put forward. Lowe *et al.* (2015) suggest that traits commonly found in tree species make them largely resilient to deleterious consequences of fragmentation. For example, the effects of fragmentation may be alleviated through extensive gene flow (Bacles et al. 2005; O'Connell et al. 2007; Davies et al. 2013), flexible mating systems (Ward et al. 2005), longevity (Piotti 2009) and overlapping generations (Petit & Hampe 2006; Bacles & Jump 2011). Another potential solution is that many populations are not yet at equilibrium (Bolliger et al. 2014), and there is a time lag (Epps

& Keyghobadi 2015) between the physical fragmentation and the genetic consequences becoming apparent in population genetic metrics (Whitlock 1992). Thus, if fragmentation of European forests is relatively recent, assessments of genetic structure in European tree species may reflect past pre-fragmentation landscapes. This delay makes it difficult to assess how contemporary anthropogenic changes in population size, structure and connectivity are truly affecting the long term genetic viability of populations. The challenge for conservation genetics is therefore to disentangle contemporary from historical evolutionary processes before irreversible fitness declines set in. This is of particular importance in retreating range edge populations that may harbor important genetic diversity (Hampe & Petit 2005).

A meta-analysis of studies on woody plants by Vranckx et al. (2012) found consistent negative genetic consequences of fragmentation, based on studies employing amplified fragment length polymorphism, random amplified polymorphic DNA and simple sequence repeat markers. Such markers typically have high mutation rates, so they are expected to approach mutation-drift-migration equilibrium more quickly than other loci (Nichols & Freeman 2004). Since 2012, most studies on population genetics in non-model organisms have used either a small number of multi-allelic SSR markers genotypes in large population samples (e.g. Mattioni et al. 2013; Davies et al. 2013; Mandák et al. 2016), or large numbers of bi-allelic SNP markers genotyped in smaller samples (e.g. Hamlin & Arnold 2014; Manthey & Moyle 2015; Cavender-Bares et al. 2015) which have different mutational dynamics, giving different patterns of polymorphism (Payseur & Cutter 2006).

We developed a study system with three key characteristics for investigating these issues. Firstly, it concerns a woody study species with both fragmented and unfragmented populations, and similar past colonization history. Secondly, the fragmentation severity varies in space and, by implication, time in the recent past. Thirdly, we could survey a suite of neutral genetic markers that evolve at different rates and thus are sensitive to the various timeframes in which the genetic consequences of fragmentation may be occurring.

We focus on dwarf birch (*Betula nana*), a monoecious, wind-pollinated subarctic dwarf tree, with a circumpolar distribution (Ashburner & McAllister 2013). Pollen records and historical evidence indicate that this species has declined significantly in the UK, attributed to changing climate and land management over past centuries (Aston 1984; Wang et al. 2014; Gilbert & Di Cosmo 2009). Its UK distribution is characterized by a northward retreat with greater population fragmentation and reduction on the southern range periphery. It is now under active conservation management on some Highland estates. By contrast dwarf birch has a different distribution in much of the subarctic, including large tracts of northern Scandinavia, where it is a keystone species of the boreal forests, with largely continuous populations. Both regions are thought to have a similar colonization history since the last glacial maximum (LGM) (Alsos et al. 2007; Dąbrowska et al. 2006), though we do not know if there was a bottleneck upon colonization of the UK after the LGM causing low diversity among UK populations. Thus, if UK populations are genetically depauperate, this could be due mainly to recent population decline and fragmentation, or alternatively be a legacy of the original colonization process. This study uses genetic data to try and distinguish these possiblities.

We develop four genomic marker sets that differ in mutation mechanism and rate: conventional microsatellites amplified with PCR primers (for clarity termed here PCR-SSRs), single nucleotide polymorphisms from RAD-sequencing reads divided into transitions (RAD-SNP_{ti}) and transversions (RAD-SNP_{tv}), and microsatellites data-mined from RAD-sequencing reads (RAD-SSRs). Genomes contain vastly more SNPs than SSRs, whilst SSRs normally display substantially more variation than individual SNPs (Payseur & Cutter 2006). Mutation rate is also considered to be substantially higher in SSRs than in SNPs (Li et al. 2002). For example, studies in *Zea mays* reported mean SSR mutation rates of 7.7×10^{-4} per locus per generation (Vigouroux et al. 2002), in *Triticum turgidum* 2.4×10^{-4} (Thuillet 2002), in *Cicer arietinum* 1×10^{-2} – 3.9×10^{-3} (Udupa & Baum 2001) and in *Pinus sylvestris* 2.4×10^{-4} (Kuchama et al. 2011). By contrast, the rate of spontaneous SNP mutations in plants are several orders of magnitude

lower. For example, in *A. thaliana* the SNP mutation rate is estimated at $7x10^{-9}$ per locus per generation (Ossowski et al. 2009).

Different categories of SNPs mutations appear to have different rates. Among substitutions and segregating polymorphism transitions are found to be substantially more abundant than transversions (Van Bers et al. 2010; Zhang & Gerstein 2003), despite the fact that eight nucleotide mutations events can lead to a transversion but only four can lead to a transition. In protein coding regions, transitions are less likely to result in amino acid substitutions, and thus more likely to persist as SNPs (Goldman & Yang 1994). Because of these differences in effective mutation rate, we divide the RAD-SNPs into transitions (RAD-SNP_{ti}) and transversions (RAD-SNP_{tv}) and compare population statistics. We also compare population genetic statistics for shorter RAD-SSRs with longer PCR-SSRs, since the apparent mutation rate may be affected by the different allele-lengths and the different ascertainment processes for the two categories. Marker loci with higher mutation rates have the advantage that they are expected to reach mutation-drift-migration equilibrium more quickly, thus permitting the effects of migration and mutation to be distinguished sooner. However, unless treated appropriately (*c.f.* Nichols & Freeman 2004) they can suffer the disadvantage of greater homoplasy potentially resulting in overestimates of gene flow.

In this study, we have four specific aims. First, to evaluate the relative inferences possible from SSR and SNP based markers in light of their differing mutation rates in non-equilibrium populations and provide a framework for analysis. Second, to characterize the genetic diversity and structure of dwarf birch across fragmented and non-fragmented landscapes. Third, to test whether the current distribution of genetic diversity is a result of recent fragmentation events or early patterns of colonization after the last glacial maximum. Fourth, to evaluate the effective population size at which the balance between genetic drift gene flow tips in favour of divergence (driven by drift). Finally, we discuss our findings in the context of conserving genetic diversity in fragmented plant populations and the choice of markers in such studies.

3.2 Materials and Methods

A total of 1220 *B. nana* individuals were collected from 29 populations across the species' extant range in the UK and 10 populations from a central portion of the species' European distribution in Scandinavia, during the summers of 2012 – 2014 (Table 1). Total genomic DNA was extracted from dried cambial tissue following modified cetyltrimethylammonium bromide (CTAB) protocol published in (Wang et al. 2013) and based on the approach of Doyle & Doyle (1987). The resulting DNA pellets were resuspended in 50-200ul of TE buffer to normalize concentrations. Quality of samples was assessed using 1% gel electrophoresis and a Nanovue spectrophotometer (GE Healthcare, UK). Approximately 9% of samples were rejected due to degraded DNA, found mainly in UK samples where small plant size limited the amount of tissue available. In all, 1115 sampled were retained from 39 populations. Nuclear microsatellite simple sequence repeats (PCR-SSRs) were amplified in all retained individuals and RAD sequencing was conducted in a subset of approximately 17% of individuals. Where samples were selected for RADseq, an additional 70% and 95% ethanol wash was added as a final stage before desiccation and rehydration in TE buffer. Furthermore, to achieve the necessary yield of 1ug high molecular weight DNA, multiple extractions were used for some individuals and the samples pooled.

3.2.1 PCR-SSR development and genotyping

This study utilises both previously published PCR-SSR loci (Kulju et al. 2004; Truong et al. 2005) and PCR-SSR loci developed here *de novo* specifically for *B. nana* (Table S1) using the v4 Dwarf birch genomic resource (http://birchgenome.org/) (Wang et al. 2013). De novo dinucleotide SSR motifs were identified using a local SequenceServer BLAST (Priyam et al. 2015). Scaffolds with high hit rates were identified and then transferred into the software WEBSAT (Martins et al. 2009) to enable the identification of repeats within the scaffold and the design of primer flanking sequences. Only one PCR-SSR assay was developed per scaffold in an effort to minimise linkage of markers.

A total of 24 PCR-SSR markers were selected and tested for stutter peaks and repeatability in a subset of 12 *B. nana* individuals. Combinations of useable loci were then tested *in silico* using the software AutoDimer (Vallone & Butler 2004) to identify possible primer dimers or cross amplification. Suitable multiplex panels (Table S1) were run on a 2100 Bioanalyzer (Agilent Technologies, CA, USA) to confirm expected product sizes. Fragment length was determined by auto capillary gel electrophoresis on an ABI 3730xl (Applied Biosystems). Trace peaks were called manually using the software GeneMarker® (Softgenetics, LCC) using default quality control settings. The six plastid loci were found to be largely monomorphic (>0.99) and were excluded from further analysis.

The PCR-SSR data were checked for presence of null alleles, deviation from Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) using the software GENEPOP v4 (Rousset 2008). LD and HWE were adjusted for multiple comparisons using a Bonferonni correction. Loci were tested for signatures of selection using the method of Beaumont and Nichols (1996) implemented in the software Lositan (Antao et al. 2008). Lositan was run initially with all loci over 50,000 iterations to estimate mean neutral F_{ST}. Then putative selected loci (>99% CI) were removed and the mean neutral F_{ST} re-estimated. A final run with all loci, using the re-estimated value for F_{ST} reduces the bias on reported values for selection. Putative clones were identified in the software Poppr (Kamvar et al. 2014).

3.2.2 RAD sequencing, marker discovery and genotyping

A subset of 190 individuals from 36 populations were selected for RAD sequencing. We elected to use the cleavage enzyme *PstI*, as previous analysis of the draft *B. nana* genome identified 70,954 cut sites (Wang et al. 2013). The resulting DNA was normalized and submitted to FLORAGENEX (Portland, Oregon) for generation and sequencing of RAD tags following the protocol of Baird et al. (2008) and Hohenlohe et al. (2010). In brief, libraries were prepared with 96 unique 8-base barcodes, ligated via adaptors to *PstI* digested genomic DNA. The resulting fragments were sequenced in two lanes of an Illumina GAIIx platform with single-end

1x100bp chemistry. A single internal control sample generated from *Saccharomyces bayanus* was included in each lane.

Raw libraries were demultiplexed and barcode sequences checked for errors using the *process_radtags* module of Stacks v1.35 (Catchen et al. 2011; Catchen et al. 2013). Reads were aligned to the *B. nana* genome and adaptor sites trimmed using the software Bowtie2 (Langmead & Salzberg 2012). Three individuals were discarded due to low coverage. SNP calling was performed using a bounded model in the *ref_map.pl* pipeline of Stacks v1.35, with two nucleotide mismatches permitted (-n 2), identified as an optimum threshold in Ilut et al. (2014) and a minimum stack depth of five (-m 5). Calls were corrected using the *rxstacks* module with a minimum log likelihood filter of -10 to retain a catalog locus (--lnl_lim -10) and other parameters set to default. Subsequently, *cstacks* and *sstacks* were manually rerun to rebuild and match to the catalog.

The *populations* program was used to filter the called SNPs based on a number of criteria. SNPs were retained that were present in a minimum of 70% of individuals in either of the sampling regions separately, or 70% of individuals across both regions combined. SNPs that significantly deviated from Hardy-Weinberg equilibrium in >5 populations (excluding populations with 1-2 individuals) were removed. The final seven bases of each locus were omitted, because these showed a relatively high frequency of SNPs possibly due to sequencing error. Putative SNPs from base 28 in all RAD loci were also removed, as the number present were unusually elevated indicating a possible error in one of the plate cycles.

For downstream analyses (except private alleles), polymorphisms with a minor allele frequency (MAF) of <0.01 were excluded unless indicated otherwise. Similarly, to reduce linkage disequilibrium for Bayesian clustering analyses, the dataset was filtered to retain only a single locus per scaffold or contig in the *B. nana* genome assembly. This strategy was considered to retain more information than if a single SNP had been retained per locus (e.g. using --write_single_snp). Finally, SNPs were exported in formats compatible with other

analysis software using custom scripts and PGDSpider 2.1.0.1 (Lischer & Excoffier 2012). For subsequent analyses we divided SNP loci into transition (RAD-SNP_{ti}) and transversion (RAD-SNP_{tv}) datasets based on the type of polymorphism recorded.

3.2.3 RAD-SSR marker development

To generate RAD-SSR markers we filtered raw Illumina RADseq reads for SSR motifs by searching the consensus catalog for loci with di- and tri- nucleotide repeats present in >50 individuals. Reads were trimmed in the same manner as previously described and the frequency of each sequence string calculated (at least three identical reads were required for inclusion, and a minimum of 10 reads per individual). Where a single sequence was present in an individual it was considered to be homozygous, where a second sequence was recorded in more than 20% of reads then it was called as a heterozygote (see Data Accessibility for script details). Alleles were designated integers values (unrelated to allele length or frequency).

3.2.4 Characterizing genetic diversity

Comparative statistics across all marker sets were calculated in Hierfstat (Goudet 2005) and Adegenet 1.4.2 (Jombart 2008), including observed heterozygosity (H_o) and expected heterozygosity (H_e) within populations. Gene diversity among samples (D_{st}) was also calculated, and combined with H_e to give overall gene diversity (H_t). Finally we calculate global measures of inbreeding (F_{is}) and (D_{est}). All calculations were performed in R (R Development Core Team 2014) and RStudio (RStudio Team 2015).

For all PCR-SSR, RAD-SSR and RAD-SNP datasets, population based allelic richness (A_r) was estimated using Hierfstat and Adegenet rarefied to a single diploid individual. Rarefied private allele frequency (P) was estimated in ADZE (Szpiech et al. 2008) to allow correction for uneven sample size. Population based H_o and H_e were calculated in GenAlEx. F_{is} was calculated using custom scripts using the formula $F_{is} = \pi - H_o/\pi$ as described in Catchen et al. (2013).

To assess variation in genetic diversity across the study area, H_e, P and F_{is} estimates were plotted against population latitude and trends assessed using linear regression. Allele frequency spectrum plots were produced for each marker and region, to provide inference about past population history. To assess evidence of genetic structure in *B. nana* populations, Bayesian clustering was performed on each dataset in the software STRUCTURE (Pritchard et al. 2000). STRUCTURE analysis was performed with the number of pre-assigned genetic clusters (K) ranging from one to ten, no-admixture and correlated allele frequencies assumed due to anticipated recent joint ancestry. We used the LOCPRIOR model (Hubisz et al. 2009) to improve accuracy by including population sampling locations as prior information. Following the advice of Pritchard in Benestan et al. (2016) we opted to divide computing time over many independent shorter runs, rather than few long runs. This is because STRUCTURE converges fairly quickly, but is poor at exploring alternative peaks in parameter space. Thus 50 independent runs were performed for each K, with 20,000 Markov chain Monte Carlo repetitions after 20,000 burn-in iterations. Longer runs were performed initially and the behavior of the MCMC checked using the *acf* function in R. The posterior probability of K was estimated using the statistic ΔK (Evanno et al. 2005) implemented in Structure Harvester (Earl & vonHoldt 2012) and visualized using CLUMPAK (Kopelman et al. 2015).

3.2.5 Demographic history

To infer the evolutionary and demographic history of dwarf birch in the UK since colonization, we used coalescent-based approximate Bayesian computation (ABC) implemented in DIY-ABC v2.0 to compare alternative scenarios (Cornuet et al. 2014). Specifically we sought to test whether contemporary observed patterns in genetic markers better support recent divergence of UK populations, or whether differentiation was established at colonization (see Ibrahim et al. 1996) and maintained by a large N_e. We also tested whether the data shows evidence of a historic bottleneck immediately on colonization or multiple bottlenecks after population differentiation, potentially as a result of recent anthropogenic impacts (see Figure 3.1 for

scenario details). Whilst there are a substantial number of possible colonization scenarios, we limited our analysis to two biologically plausible models with six variables, as overly complex scenarios can result in poor parameter estimates (Bertorelle et al. 2010). To reduce the complexity of the analysis further we subsampled our full dataset, retaining 12 representative well-spaced populations from across the UK distribution, and combining populations RA/RW, and MU1/MU2 on the basis of their close geographical proximity. All markers were filtered for <25% missing data and genotypes in every population, and a minor allele frequency >0.05. In each case this reduced dataset was tested together with an out-group of three pooled Scandinavian populations from the center of our sampling distribution: populations KP/KR/SM.

We generated 10^6 simulations for each scenario across each dataset. We used preliminary runs with very broad priors together and incremental prior checking steps, together with our knowledge of the species' ecology to inform the range of our priors. We limited contemporary N_e to 10-2000 individuals in the UK; historic UK and Scandinavian N_e to 10-10,000; and a maximum time for all divergence and bottleneck events of 3000 generations before present, all with uniform prior distributions. Full summary statistics and prior parameters are reported in Tables 3.1, 3.2, and S2.

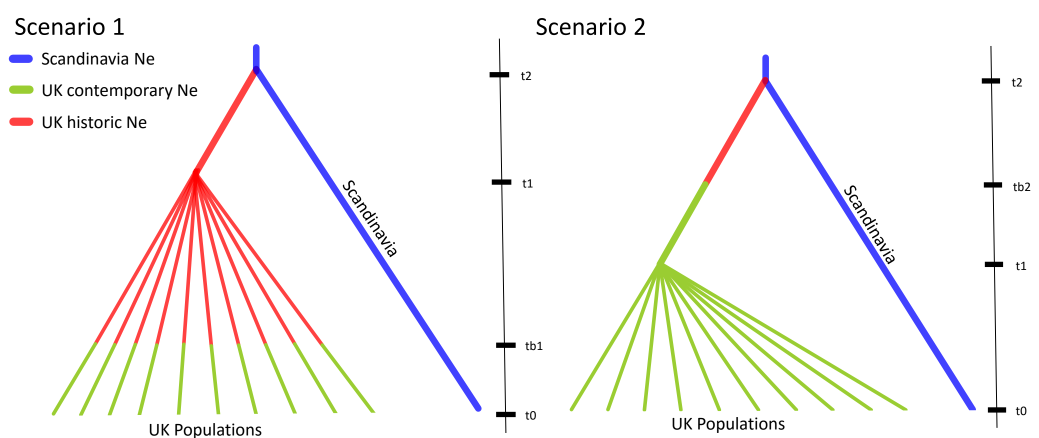


Figure 3.1 Diagram of proposed scenarios to describe the evolutionary history of dwarf birch after the last glacial maximum. The key difference is that scenario A implies multiple more recent bottlenecks after differentiation of populations. Scenario B implies a bottleneck before population differentiation most likely as a result of a limited number of colonizing individuals.

To evaluate our results, we checked that our observed data fell within the range of simulated datasets using principal component analysis. Once satisfied, the posterior probabilities of scenarios were estimated using logistic regression method on the 1% of simulated datasets closest to the observed data. Posterior parameter estimation was performed for the best supported scenario. In the literature, SSRs have been reported as being less suited to questions of deeper evolutionary timescales on account of homoplasy, whereas SNPs may lack the variation to distinguish recent demographic events (Morin et al. 2004). In this analysis we achieved greater resolution and confidence in our results by combining information from the posterior estimates of each dataset into a single density distribution.

Table 3.1. Prior parameters used in DIY-ABC analysis. Population size units are number of individuals. Time units are in generations**.**

Variable	Description	Minimum	Maximum
NeS	Current UK effective population size	10	1000
NeS2	Historic UK effective population size	10	10000
NeF	Current Scandinavian effective population size	10	10000
tb1	Putative contemporary bottleneck	10	3000
tb2	Putative historic bottleneck	10	3000
t1	Divergence of UK regional populations	10	3000
t2	Divergence of Scandinavian and UK linneages	10	3000

Defined priors: All uniform distribution; t2>t1, t2>tb2, t2>tb1, tb2>tb1, t1>tb1.

Table 3.2 Summary statistics used in DIY-ABC analysis.

Summary Statistic	PCR-SSR	NGS-SSR	NGS-SNP_{ti}	NGS-SNP_{tv}
Individual based				
Mean number of alleles	x	x		
Mean genic diversity	x	x	x	x
Variance of genic diversity			x	x
Mean size variance	x			
Mean Garza-Williamson's M	x			
Population based				
Mean number of alleles	x	x		
Mean genic diversity	x	x		
Mean size variance	x			
F_{ST}	x	x	x	x
Variance F_{ST}			x	x
Shared allele distance	x	x		
$(du)_2$ distance	x			

3.2.6 Population differentiation and genetic drift

Two different approaches were employed to assess population differentiation. Pairwise F_{ST} between all population combinations was computed in Arlequin v3.5.2 (Lischer 2010) with significance tested over 10,000 iterations. Geographic distance matrices were calculated from population grid references using Geographic Distance Matrix Generator V1.2.3 (Ersts, accessed 2016-02-11). Evidence for isolation by distance and regional discontinuities in genetic diversity were assessed using Mantel tests within fragmented and non-fragmented regions, performed in *vegan* (Dixon 2003). To visualize the relationship, pairwise F_{ST} was transformed using the formula $M = (1- F_{ST})/F_{ST}$ and plotted against log transformed geographic distance.

In addition, to test the divergence of each population from the regional mean a maximum likelihood estimate of F_{ST} was calculated for each population against the combined regional allele frequency for each locus using custom R scripts (see Data Accessibility). This calculation was conducted separately for the UK and Scandinavia and the F_{ST} estimate plotted against a linkage-disequilibrium based measure of effective population size calculated by the program NeEstimator v2 (Do et al. 2014), using PCR-SSR markers (chosen due to their larger sample size), and against rarefied allelic richness.

3.3 Results

The location and sample sizes of surveyed *B. nana* populations are reported in in Table 3.3 (also see Figure 2.2 for reference). After quality control of DNA extractions, 1115 individuals were retained for PCR-SSR and 187 individuals for RADseq analyses (Table 3.3). Forty-nine individuals from the PCR-SSR dataset (4.4%) were subsequently identified as clonal genotypes and removed, the majority of which were in the UK.

3.3.1 PCR-SSR marker validation

The number of PCR-SSR alleles per nuclear locus ranged from two to 42, with 446 alleles in total (Table S1). Four loci had an estimated null allele frequency greater than 0.1. The same loci were found to deviate consistently from Hardy-Weinberg equilibrium (HWE) across populations, and outlier tests indicated two (L3.1 and NU3B) may have been under positive selection. Thus all four were excluded from analysis of genetic differentiation. Six populations were found to significantly deviate from HWE at >2 of the remaining loci, these were AS, FS, LG, LL, MD and RW, all in Scotland. Significant linkage disequilibrium (LD) was detected in 237 of 6120 pairwise comparisons, however no consistent locus-locus pattern could be identified. Almost all (219) cases of significant LD were detected in just six of the 39 populations; LG, LL, LR, MD, MO and PC, indicating whilst loci may not be closely linked, these six populations may have small effective size, for example due to long term reduction in population size or past population bottlenecks.

Table 3.3 *Betula nana* sampling locations, altitude and sample size across the UK and Scandinavia. PCR gives sample sizes for PCR-SSR data. RAD gives sample sizes for RAD-SSR, RAD-SNP$_{ti}$ and RAD-SNP$_{tv}$ datasets. Values in parenthesis are the number of putative clones.

Region	Population	ID	Latitude	Longitude	Altitude	PCR	RAD
	Nordkapp	NK	70.863	25.722	10	30 (0)	6
	Sirbma	SM	70.052	27.561	55	30 (4)	6
	Tenontie	TE	69.944	26.722	113	30 (0)	6
	Skalluvaara	SK	69.798	27.080	217	32 (0)	6
Scandinavia	Kevo Plateau	KP	69.773	26.955	318	32 (0)	6
	Kevojarvi	KA	69.763	26.981	124	31 (2)	6
	Gearddosjavri	UG	69.732	26.937	121	23 (2)	6
	Kevo Reserve	KR	69.671	26.960	218	31 (0)	6
	Kotilampi	DJ	69.313	26.653	216	29 (0)	6
	Partakko	PT	69.272	27.988	124	28 (0)	6
				Sub-Total		296 (8)	60
	Ben Loyal	BL	58.401	-4.404	300	30 (0)	6
	Meall Odhar	MO	58.163	-4.423	404	25 (11)	6
	Beinn Enaiglair	BE	57.786	-5.009	480	27 (0)	5
	Luichart	LH	57.725	-4.900	268	32 (2)	6
	Ben Wyvis	BW	57.650	-4.602	482	34 (0)	5
	DJG Ben Wyvis	DG	57.646	-4.556	472	21 (0)	-
	Loch Meig	ME	57.534	-4.804	450	26 (0)	6
	Glen Cannich	GC	57.345	-4.856	455	33 (0)	6
	Faskanyle	FS	57.327	-4.848	486	31 (0)	-
	Dundreggan	DE	57.231	-4.754	448	30 (0)	6
	An Suidhe	AS	57.224	-4.812	661	30 (2)	2
	Beinn Bhreac	BB	57.211	-4.823	500	33 (0)	6
	Portclair	PC	57.204	-4.639	478	41 (5)	6
UK	River Avon	AV	57.137	-3.491	549	31 (0)	6
	Monadhliaths	MD	57.058	-4.307	712	33 (5)	6
	Meall an tslugain	SL	57.045	-3.451	633	31 (0)	6
	Loch Muick 1	MU1	56.920	-3.198	492	31 (0)	6
	Loch Muick 2	MU2	56.918	-3.205	517	32 (0)	6
	Loch Laggan	LG	56.889	-4.545	364	49 (6)	6
	Loch Loch	LL	56.846	-3.647	673	30 (4)	6
	Ben Gullabin	BG	56.840	-3.467	594	5 (0)	1
	Loch Rannoch	LR	56.758	-4.415	499	29 (2)	6
	Rannoch West	RW	56.650	-4.785	306	31 (2)	6
	Rannoch Moor B	RB	56.603	-4.740	304	31 (2)	6
	Rannoch Moor A	RA	56.603	-4.738	295	30 (0)	-
	Lennox	LX	55.970	-4.276	164	9 (0)	2
	Emblehope	EM	55.244	-2.483	448	2 (0)	1
	Spadeadam	SA	55.053	-2.568	275	1 (0)	1
	Teesdale	TD	54.654	-2.280	499	2 (0)	2
					Sub-Total	770 (41)	127
					Total	1066	187

3.3.2 RAD-SNP discovery

A total of 339,291,129 reads were generated from two single-end Illumina HiSeq lanes, of which 81.66% passed strict quality controls. Three individuals were removed due to low sequencing coverage. The number of reads retained per individuals ranged from 373,727 to

2,852,627, with a mean of 1,397,192. Across all individuals an average of 78.3% of reads aligned to the *B. nana* reference genome. Alignment was consistently higher in Scandinavian samples, despite the reference genome being of Scottish provenance. We attribute this to poorer quality DNA derived from extractions on Scottish samples, due to the smaller size of twigs and difficulties in removing all bark from cambium tissue. Despite approximately 20% of reads that did not align being excluded, the reference-mapped pipeline identified approximately a third more usable loci than when the analysis was undertaken de novo (data not shown). This is perhaps due to reference based mapping performing better at distinguishing paralogs. A threshold of 70% presence resulted in 6,092 loci containing 23,882 SNPs. Quality filtering retained 17,694 SNPs with an overall Transition/Transversion ratio of (11:10). With minor alleles occurring at a frequency of <0.01 removed, 4,208 loci and 8,081 SNPs were retained, of which 4,775 were transition and 3,306 transversion polymorphisms (144:100); thus proportionately more of the low frequency polymorphisms were transitions.

3.3.3 RAD-SSR discovery

Simple sequence repeat motifs were identified in 3,507 RADseq loci. After filtering for loci present in >50% of sampled individuals, 197 loci were retained. We then filtered for variable loci by rejecting those with very high major allele frequency (as is typical in the development of PCR-SSR) markers resulting in 38 retained markers. Finally we excluded four loci with more than 30 unique genotypes due to the risk of a high number of sequencing errors in these reads. The number of alleles per locus ranged from 3 to 29 with a total of 354. Outlier analysis identified seven loci under putative selection, which were excluded from analysis of differentiation. No consistent pattern of deviation from HWE or LD was detected.

3.3.4 Marker comparison

The number of alleles per dataset ranged from 354 in RAD-SSRs to 9550 in RAD-SNP_{ti}s (Table 3.4). Mean private allele frequency (P) is also highest in PCR-SSR markers, intermediate in RAD-SSRs and lowest in RAD-SNP markers. Across all datasets, patterns of observed heterozygosity

(H_o), overall gene diversity (H_t), both within samples (D_{st}) and populations (H_e), and genetic differentiation (D_{est}) show an increase with mutation rate (Table 3.4). Inbreeding estimates (F_{is}) did not vary with mutation rate; rather they were substantially lower in the PCR-SSR dataset than in all three RAD derived datasets.

Table 3.4 Summary statistics for the four different marker sets used in this study, including sample sizes, the number of loci, alleles alleles (nAll), with alleles per locus in parenthesis, private allele frequency (P), Observed heterozygosity (H_o), expected heterozygosity (H_e), individual gene diversity (D_{st}), Overall gene diversity (H_t), inbreeding (F_{is}) and a measure of population differentiation (D_{est}).

Marker	Ind.	Pop	Loci	nAll	P	H_o	H_e	D_{st}	H_t	F_{is}	D_{est}
PCR-SSR	1115	39	18	421 (23.4)	0.114	0.552	0.677	0.066	0.743	0.185	0.209
RAD-SSR	187	36	34	354 (10.4)	0.092	0.245	0.617	0.035	0.652	0.603	0.095
RAD-SNP_{ti}	187	36	4775	9550 (2)	0.008	0.038	0.114	0.009	0.123	0.665	0.011
RAD-SNP_{tv}	187	36	3306	6612 (2)	0.008	0.037	0.111	0.008	0.119	0.665	0.009

3.3.5 Genetic diversity and population structure

Comparison of population genetic diversity statistics between fragmented and non-fragmented regions shows allelic richness, private allele frequency, observed and expected heterozygosity were significantly higher in Scandinavia across all markers (Two sample T-test, all $p<0.025$), with the exception of PCR-SSR based expected heterozygosity ($t = -0.465$, $p=0.64$) and RAD-SSR based private allele frequency ($t = 1.167$, $p=0.25$) (Table 3.5). This pattern is significant even with the exclusion of five genetically depauperate southern UK populations that show the greatest decline, but have the smallest sample sizes (data not shown). Estimates of fixation index were on average marginally higher in Scandinavia, although there was no significant difference in any dataset.

Across populations, allelic richness was highest in PCR-SSR markers and lowest in the RAD-SNP_{tv} markers, but variation was highest in the RAD-SSR dataset, followed by PCR-SSRs and then RAD-SNP markers (Table 3.5). Calculations of observed and expected heterozygosity showed that almost all populations displayed heterozygote deficit. To assess evidence of

genetic decline across a retreating range in the UK, gene diversity, fixation index and private allele frequency were regressed against latitude and compared to a central portion of the range in Scandinavia: increasing expected heterozygosity was significantly predicted by increasing latitude in the UK across all markers (Figure 3.2; PCR-SSR: $F_{1,27} = 11.13$, $P = 0.002$; RAD-SSR: $F_{1,24} = 8.57$, $P = 0.007$; RAD-SNP_{ti}: $F_{1,24} = 23.07$, $P = <0.001$; RAD-SNP_{tv}: $F_{1,24} = 21.13$, $P = <0.001$), no such relationship was found in Scandinavia. Unexpectedly, the Fixation index also increased significantly with latitude in Scotland, (Figure 3.2; PCR-SSR: $F_{1,26} = 23.45$, $P = <0.001$; RAD-SSR: $F_{1,21} = 9.726$, $P = 0.005$; RAD-SNP_{ti}: $F_{1,24} = 6.555$, $P = 0.018$; RAD-SNP_{tv}: $F_{1,24} = 6.558$, $P = 0.018$), there was no significant pattern in Scandinavia. Only RAD-SNP_{tv} show a significant pattern of increasing private allele frequency with latitude in Scotland ($F_{1,24} = 13.51$, $P = 0.001$), all other relationships were non-significant. Allele frequency spectrum plots are presented in Figure S5.

In STRUCTURE analyses, the best-supported value of K was K = 2 for all markers sets. In the RAD-SNP datasets, the UK and Scandinavian regions were highly differentiated with confident correct assignment of all but one individual from Teesdale in the RAD-SNP_{tv} dataset (Figure 3.3). The PCR-SSR dataset largely discriminated between the UK and Scandinavia, except for a scattering of individuals, perhaps due to homoplasy. RAD-SSRs displayed much poorer cluster assignment with very few individuals being assigned to clusters with 100% support.

Table 3.5 Summary statistics for all populations ordered by decreasing latitude. Allelic richness (A_r) and private alleles (P) are rarefied to a single diploid individual.

Pop	A_r				P				H_o				H_e				F_{is}			
	PCR-SSR	RAD-SSR	RAD-SNP_{ti}	RAD-SNP_{tv}	PCR-SSR	RAD-SSR	RAD-SNP_{ti}	RAD-SNP_{tv}	PCR-SSR	RAD-SSR	RAD-SNP_{ti}	RAD-SNP_{tv}	PCR-SSR	RAD-SSR	RAD-SNP_{ti}	RAD-SNP_{tv}	PCR-SSR	RAD-SSR	RAD-SNP_{ti}	RAD-SNP_{tv}
NK	1.722	1.538	1.111	1.105	0.132	0.024	0.012	0.010	0.448	0.215	0.043	0.043	0.591	0.464	0.098	0.099	0.264	0.613	0.525	0.552
SM	1.737	1.651	1.120	1.109	0.183	0.221	0.012	0.009	0.503	0.275	0.047	0.046	0.603	0.564	0.105	0.106	0.194	0.582	0.522	0.563
TE	1.716	1.598	1.108	1.104	0.129	0.210	0.016	0.014	0.557	0.222	0.033	0.034	0.702	0.504	0.092	0.096	0.211	0.641	0.613	0.639
SK	1.687	1.576	1.114	1.103	0.223	0.056	0.011	0.012	0.481	0.278	0.040	0.041	0.552	0.493	0.100	0.102	0.154	0.517	0.552	0.558
KP	1.696	1.577	1.117	1.112	0.160	0.085	0.013	0.015	0.454	0.294	0.051	0.050	0.569	0.507	0.105	0.106	0.202	0.488	0.470	0.473
KA	1.666	1.546	1.104	1.100	0.099	0.073	0.009	0.011	0.451	0.274	0.044	0.044	0.545	0.470	0.092	0.095	0.167	0.469	0.516	0.509
UG	1.755	1.564	1.113	1.108	0.292	0.110	0.013	0.008	0.487	0.365	0.056	0.056	0.604	0.505	0.101	0.100	0.224	0.376	0.407	0.413
KR	1.705	1.569	1.116	1.109	0.142	0.144	0.011	0.013	0.452	0.282	0.052	0.052	0.577	0.508	0.103	0.105	0.240	0.469	0.477	0.454
DJ	1.683	1.543	1.117	1.108	0.138	0.137	0.012	0.010	0.461	0.273	0.053	0.054	0.559	0.475	0.104	0.106	0.220	0.427	0.468	0.500
PT	1.727	1.538	1.116	1.114	0.155	0.071	0.011	0.009	0.497	0.295	0.049	0.050	0.592	0.479	0.103	0.105	0.193	0.443	0.478	0.497
Mean	1.710	1.570	1.114	1.107	0.165	0.113	0.012	0.011	0.479	0.277	0.047	0.047	0.589	0.497	0.100	0.102	0.207	0.503	0.503	0.516
BL	1.676	1.527	1.091	1.086	0.104	0.136	0.005	0.009	0.529	0.165	0.029	0.030	0.662	0.454	0.075	0.074	0.202	0.722	0.581	0.622
MO	1.659	1.431	1.102	1.097	0.052	0.057	0.007	0.010	0.604	0.150	0.030	0.029	0.645	0.369	0.085	0.085	0.071	0.627	0.643	0.642
BE	1.699	1.488	1.083	1.083	0.156	0.056	0.004	0.007	0.562	0.160	0.028	0.028	0.684	0.369	0.065	0.066	0.179	0.618	0.613	0.610
LH	1.685	1.533	1.106	1.092	0.099	0.087	0.008	0.009	0.578	0.159	0.034	0.034	0.672	0.452	0.090	0.092	0.142	0.668	0.606	0.581
BW	1.714	1.450	1.094	1.084	0.092	0.070	0.007	0.007	0.575	0.140	0.035	0.036	0.700	0.365	0.075	0.075	0.196	0.626	0.546	0.606
DG	1.686	-	-	-	0.095	-	-	-	0.572	-	-	-	0.667	-	-	-	0.179	-	-	-
ME	1.724	1.481	1.102	1.103	0.131	0.051	0.008	0.008	0.563	0.206	0.037	0.037	0.697	0.420	0.088	0.090	0.218	0.553	0.565	0.599
GC	1.679	1.571	1.102	1.106	0.065	0.182	0.007	0.007	0.553	0.228	0.032	0.033	0.665	0.473	0.086	0.087	0.181	0.603	0.633	0.605
FS	1.669	-	-	-	0.112	-	-	-	0.394	-	-	-	0.545	-	-	-	0.259	-	-	-
DE	1.734	1.593	1.105	1.101	0.160	0.111	0.006	0.008	0.552	0.251	0.033	0.033	0.718	0.515	0.090	0.091	0.246	0.547	0.618	0.574
AS	1.699	1.307	1.060	1.057	0.117	0.040	0.005	0.008	0.528	0.221	0.036	0.035	0.684	0.210	0.034	0.032	0.255	0.246	0.333	0.374
BB	1.695	1.558	1.096	1.101	0.102	0.182	0.008	0.007	0.541	0.254	0.035	0.035	0.682	0.476	0.082	0.080	0.244	0.512	0.541	0.501
PC	1.680	1.591	1.099	1.097	0.098	0.090	0.009	0.009	0.538	0.322	0.040	0.040	0.670	0.518	0.085	0.085	0.199	0.435	0.536	0.525
AV	1.646	1.556	1.098	1.086	0.109	0.039	0.006	0.006	0.526	0.310	0.051	0.053	0.633	0.491	0.086	0.089	0.169	0.485	0.409	0.398
MD	1.659	1.543	1.100	1.092	0.107	0.042	0.011	0.005	0.571	0.256	0.040	0.040	0.649	0.475	0.088	0.089	0.161	0.513	0.520	0.477
SL	1.695	1.622	1.100	1.093	0.058	0.047	0.007	0.006	0.605	0.311	0.038	0.038	0.683	0.539	0.086	0.085	0.147	0.496	0.553	0.553
MU1	1.602	1.446	1.096	1.090	0.108	0.060	0.007	0.004	0.515	0.226	0.034	0.034	0.591	0.392	0.083	0.085	0.162	0.442	0.601	0.560

Table 3.5 Continued.

Pop	A_r				P				H_o				H_e				F_{is}			
	PCR-SSR	RAD-SSR	RAD-SNP_{ti}	RAD-SNP_{tv}	PCR-SSR	RAD-SSR	RAD-SNP_{ti}	RAD-SNP_{tv}	PCR-SSR	RAD-SSR	RAD-SNP_{ti}	RAD-SNP_{tv}	PCR-SSR	RAD-SSR	RAD-SNP_{ti}	RAD-SNP_{tv}	PCR-SSR	RAD-SSR	RAD-SNP_{ti}	RAD-SNP_{tv}
MU2	1.569	1.549	1.103	1.087	0.076	0.014	0.005	0.004	0.465	0.243	0.041	0.042	0.559	0.485	0.091	0.091	0.168	0.523	0.516	0.518
LG	1.646	1.481	1.095	1.093	0.051	0.067	0.008	0.009	0.570	0.204	0.031	0.031	0.639	0.412	0.080	0.079	0.110	0.585	0.611	0.615
LL	1.651	1.631	1.104	1.096	0.061	0.061	0.005	0.008	0.497	0.253	0.044	0.044	0.637	0.547	0.091	0.089	0.234	0.630	0.513	0.513
BG	1.431	1.269	1.032	1.038	0.031	0.000	0.011	0.009	0.478	0.206	0.023	0.022	0.386	0.103	0.011	0.011	0.002	-	-	-
LR	1.665	1.477	1.099	1.095	0.110	0.056	0.008	0.005	0.616	0.230	0.040	0.040	0.649	0.422	0.086	0.087	0.077	0.515	0.496	0.480
RW	1.670	1.536	1.102	1.097	0.090	0.059	0.008	0.005	0.547	0.304	0.039	0.039	0.659	0.471	0.090	0.089	0.169	0.420	0.543	0.472
RB	1.630	1.573	1.094	1.091	0.091	0.098	0.008	0.007	0.580	0.259	0.040	0.041	0.619	0.503	0.082	0.083	0.092	0.546	0.488	0.564
RA	1.645	-	-	-	0.107	-	-	-	0.568	-	-	-	0.633	-	-	-	0.118	-	-	-
LX	1.566	1.227	1.048	1.057	0.075	0.033	0.006	0.005	0.354	0.132	0.019	0.019	0.442	0.103	0.028	0.028	0.262	0.257	0.522	0.616
EM	1.583	1.320	1.029	1.040	0.169	0.297	0.005	0.007	0.528	0.235	0.022	0.025	0.438	0.118	0.011	0.012	0.075	-	-	-
SA	1.389	1.389	1.042	1.036	0.063	0.000	0.008	0.002	0.389	0.206	0.031	0.031	0.194	0.103	0.016	0.015	-	-	-	-
TD	1.667	1.250	1.038	1.045	0.119	0.228	0.005	0.005	0.667	0.103	0.013	0.013	0.500	0.129	0.016	0.015	0.007	0.629	0.457	0.537
Mean	1.645	1.477	1.085	1.082	0.097	0.083	0.007	0.007	0.537	0.221	0.034	0.034	0.607	0.381	0.069	0.069	0.161	0.530	0.541	0.545

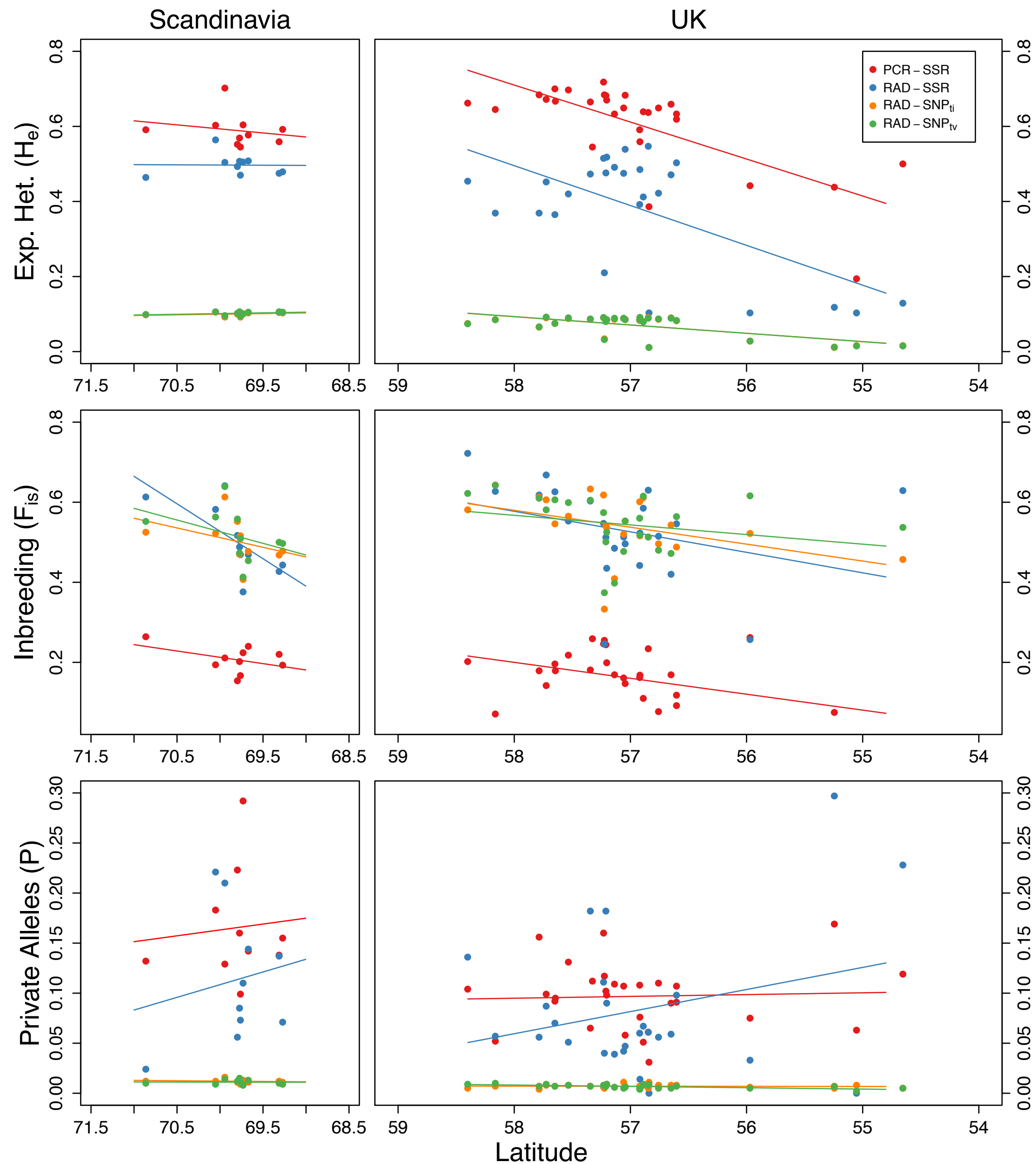


Figure 3.2 Diagnostic plots of population based expected heterozygosity, inbreeding and rarefied private alleles across four marker datasets. Linear regression trend lines are plotted for each marker type in each region. Expected heterozygosity is largely concordant for RAD-SNP_{ti} and RAD-SNP_{tv} datasets, thus they overlap considerably.

3.3.6 Demographic history

Comparison of alternative scenarios in DIY-ABC supported a scenario in which the major reduction in UK *B. nana* effective population sizes occurred independently across populations a long time after colonization of the UK and after local differentiation had arisen (Figure 3.4). Posterior estimates supported differentiation of UK and Scandinavian lineages 1080 (CI 599-

1453) generations ago, presumably shortly after colonization with the UK and Scandinavia maintaining large effective population sizes of 7440 (CI 5362-9371) and 8794 (CI 7700-9481) respectively (Table 3.6). Subsequently, UK populations appear to have become differentiated approximately 233 generations ago (CI 151-365) with a bottleneck resulting in an order of magnitude decline to approximately 109 (570-993) individuals, dated to 50 generations ago. Dating these events is difficult, as we do not have estimates of generation time in the wild. Here, we have used the colonization of the UK, which is likely to be more or less concurrent with the end of the last glaciation and divergence of the UK *B. nana* lineage and to estimate generation time at ~10 years (see discussion).

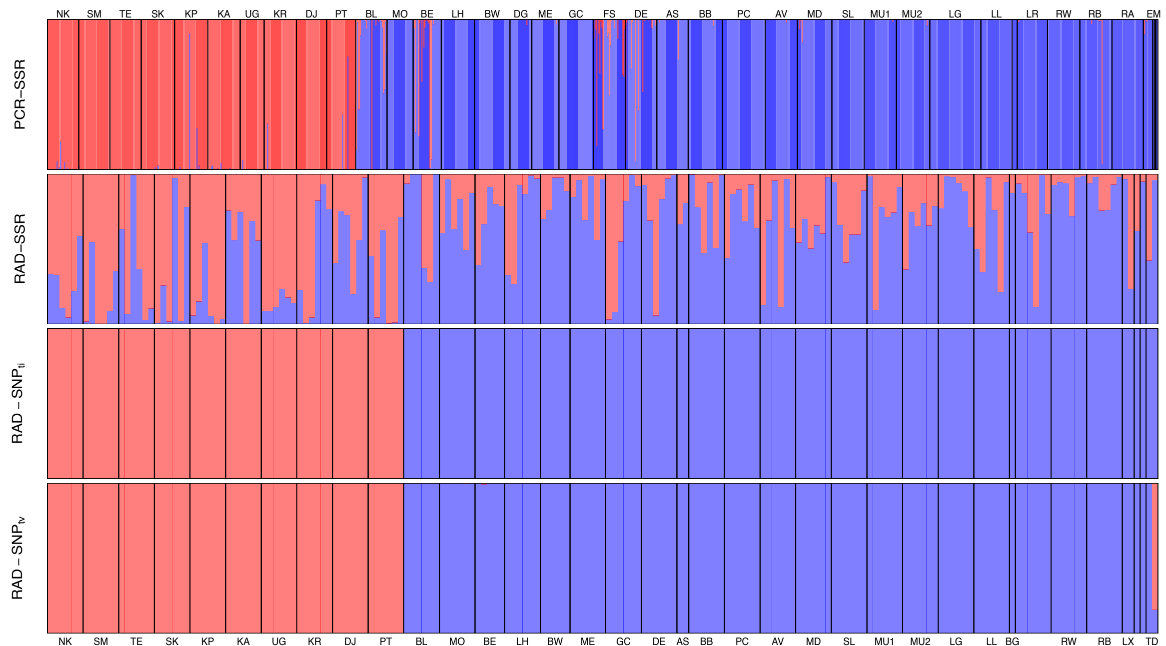


Figure 3.3 STRUCTURE plots of *Betula nana* for all marker sets with individuals ordered from left to right by decreasing latitude. The Evanno DeltaK method clearly identifies k=2 for all datasets. The PCR-SSR dataset consists of 1066 individuals (population IDs above the plot), and the other three datasets consist of 187 individuals (population IDs below the plot).

3.3.7 Genetic differentiation

Mantel tests identified significant isolation by distance (IBD) in the PCR-SSR dataset for both fragmented UK populations (r=0.28, p=0.009), and non-fragmented populations in Scandinavia

(r=0.34, p=0.044) (Figure 3.5). None of the other marker sets showed a significant pattern: RAD-SSRs (UK: r=-0.003, p=0.46; Scandinavia: r=0.36, p=0.055), RAD-SNP$_{ti}$ (UK: r=-0.003, p=0.06; Scandinavia: r=0-0.03, p=0.51) and RAD-SNP$_{tv}$ (UK: r=0.16, p=0.10; Scandinavia: r=0-0.002, p=0.45). Evidence for IBD was generally stronger in fragmented UK populations, although in all cases small *r* statistics suggest that only a very limited proportion of genetic divergence is explained by distance.

Maximum likelihood estimates of population F_{ST} are significantly higher in the UK compared to Scandinavia across all markers (t-test, all p<0.05). F_{ST} values are negatively correlated with latitude in the UK (PCR-SSR: $F_{1,27}$ = 13.91, P = <0.001; RAD-SSR: $F_{1,24}$ = 18.18, P = <0.001; RAD-SNP^{ti}: $F_{1,24}$ = 24.92, P = <0.001; RAD-SNP^{tv}: $F_{1,24}$ = 27.96, P = <0.001), but show no significant correlation in Scandinavia (Table 3.7). When plotted against linkage disequilibrium based estimates of effective population size there is a strong negative relationship for all markers (Figure 3.6). The data also clearly show a strong trend for markers with higher mutations rates to exhibit lower F_{ST} values. Linearized maximum likelihood F_{ST} values from the slowest mutation rate dataset were plotted against all other markers to compare accumulated genetic differentiation across markers (Figure S2). Regression of maximum likelihood F_{ST} against rarefied allelic richness showed a strong negative relationship confirming that genetic drift is associated with a loss of allelic diversity (Figure 3.7).

Table 3.6 Posterior demographic parameter estimates for each marker based on approximate Bayesian computation analysis. Final column indicates estimates using combined information from all marker distributions.

Para.	PCR-SSR	RAD-SSR	RAD-SNP$_{ti}$	RAD-SNP$_{tv}$	Combined
tb1	59 (20 - 229)	52 (20 - 241)	38 (16 - 103)	56 (20 - 149)	50 (24 - 94)
t1	407 (186 - 1117)	230 (98 - 690)	175 (84 - 445)	255 (114 - 609)	233 (151 - 365)
t2	807 (332 - 2595)	414 (188 - 2133)	1198 (505 - 1989)	1298 (595 - 2070)	1080 (599 - 1453)
NeS	1929 (769 - 1964)	438 (171 - 1812)	25 (16 - 366)	31 (19 - 526)	109 (570 - 993)
NeS1	6810 (2860 - 9592)	7955 (3047 - 9627)	7175 (3367 - 9647)	7959 (3883 - 9745)	7440 (5362 - 9371)
NeF	8766 (6338 - 9677)	8376 (5277 - 9426)	9414 (5305 - 9848)	9746 (6109 - 9889)	8794 (7700 - 9481)

Confidence intervals (0.05-0.95) are given in parentheses.

Table 3.7 Maximum likelihood F_{ST} across all markers and PCR-SSR based effective population size (N_e) for all populations.

Pop.	PCR-SSR (95% CI)	RAD-SSR (95% CI)	RAD-SNP$_{ti}$ (95% CI)	RAD-SNP$_{tv}$ (95%CI)	Ne (95% CI)
NK	0.01 (0.00-0.01)	0.10 (0.05-0.17)	0.08 (0.08-0.09)	0.10 (0.09-0.11)	313.9 (104.3-Inf)
SM	0.00 (0.00-0.01)	0.04 (0.02-0.09)	0.09 (0.08-0.10)	0.09 (0.08-0.10)	1915.7 (113.2-Inf)
TE	0.01 (0.00-0.02)	0.08 (0.04-0.14)	0.14 (0.13-0.15)	0.13 (0.12-0.15)	-159.9 (2583.5-Inf)
SK	0.02 (0.01-0.03)	0.07 (0.03-0.13)	0.08 (0.08-0.09)	0.09 (0.08-0.11)	227.8 (37.9-Inf)
KP	0.01 (0.00-0.01)	0.09 (0.04-0.15)	0.06 (0.06-0.07)	0.06 (0.06-0.07)	218.1 (77.6-Inf)
KA	0.01 (0.01-0.02)	0.13 (0.07-0.21)	0.14 (0.13-0.15)	0.14 (0.13-0.16)	68.1 (42.8-143.8)
UG	0.02 (0.01-0.04)	0.04 (0.01-0.09)	0.09 (0.08-0.09)	0.09 (0.09-0.10)	112.4 (22.3-Inf)
KR	0.00 (0.00-0.01)	0.05 (0.02-0.10)	0.07 (0.06-0.07)	0.07 (0.06-0.08)	12446.5 (132.4-Inf)
DJ	0.01 (0.01-0.02)	0.10 (0.05-0.17)	0.06 (0.06-0.07)	0.08 (0.07-0.09)	93.7 (45-1525.6)
PT	0.00 (0.00-0.01)	0.10 (0.05-0.17)	0.07 (0.07-0.08)	0.06 (0.06-0.07)	227.6 (99.8-Inf)
Mean	0.01	0.08	0.09	0.09	1546.39
BL	0.06 (0.04-0.08)	0.16 (0.09-0.24)	0.23 (0.22-0.25)	0.22 (0.20-0.24)	30.5 (22.3-45.1)
MO	0.08 (0.05-0.11)	0.14 (0.07-0.24)	0.18 (0.17-0.20)	0.15 (0.13-0.16)	11.4 (8.9-14.7)
BE	0.04 (0.03-0.06)	0.25 (0.15-0.37)	0.24 (0.22-0.26)	0.22 (0.20-0.25)	-428.9 (216-Inf)
LH	0.03 (0.02-0.04)	0.15 (0.09-0.22)	0.15 (0.14-0.16)	0.19 (0.17-0.21)	49.4 (35.7-75.1)
BW	0.02 (0.01-0.03)	0.14 (0.05-0.26)	0.22 (0.20-0.24)	0.21 (0.18-0.23)	99.2 (52.8-447.8)
DG	0.02 (0.01-0.03)	-	-	-	36.1 (25.1-59.1)
ME	0.03 (0.02-0.04)	0.19 (0.13-0.28)	0.18 (0.17-0.19)	0.16 (0.14-0.17)	74.2 (37-559.9)
GC	0.04 (0.03-0.06)	0.13 (0.07-0.20)	0.20 (0.18-0.21)	0.16 (0.14-0.17)	52.8 (34.4-100.3)
FS	0.04 (0.03-0.06)	-	-	-	744.8 (93.7-Inf)
DE	0.02 (0.02-0.03)	0.05 (0.02-0.10)	0.13 (0.12-0.14)	0.11 (0.10-0.12)	44.3 (28.6-85.9)
AS	0.02 (0.01-0.03)	0.29 (0.14-0.48)	0.41 (0.37-0.45)	0.39 (0.34-0.44)	85.7 (43.6-528.9)
BB	0.03 (0.02-0.05)	0.12 (0.07-0.19)	0.20 (0.18-0.21)	0.16 (0.15-0.18)	30.6 (23.6-41.4)
PC	0.05 (0.04-0.07)	0.12 (0.07-0.18)	0.19 (0.18-0.21)	0.18 (0.16-0.19)	14.8 (12.4-17.8)
AV	0.10 (0.08-0.14)	0.07 (0.04-0.13)	0.15 (0.14-0.16)	0.17 (0.16-0.19)	62.1 (35.6-172)
MD	0.09 (0.07-0.12)	0.18 (0.12-0.26)	0.19 (0.18-0.20)	0.21 (0.19-0.22)	10.8 (9.1-12.7)
SL	0.03 (0.02-0.05)	0.04 (0.02-0.09)	0.20 (0.19-0.21)	0.18 (0.16-0.19)	273.2 (95.5-Inf)
MU1	0.10 (0.07-0.13)	0.16 (0.09-0.24)	0.23 (0.22-0.25)	0.20 (0.19-0.22)	591 (85.9-Inf)
MU2	0.12 (0.09-0.16)	0.18 (0.12-0.25)	0.19 (0.18-0.20)	0.20 (0.18-0.22)	714.9 (95.8-Inf)
LG	0.05 (0.04-0.07)	0.19 (0.12-0.28)	0.23 (0.22-0.25)	0.20 (0.18-0.21)	15.1 (12.6-18.1)
LL	0.11 (0.08-0.15)	0.09 (0.06-0.15)	0.19 (0.18-0.20)	0.20 (0.18-0.21)	13.4 (10.4-17.4)
BG	0.42 (0.29-0.57)	0.54 (0.25-0.78)	0.77 (0.73-0.81)	0.67 (0.61-0.73)	0.5 (0.4-0.6)
LR	0.10 (0.08-0.14)	0.22 (0.15-0.31)	0.19 (0.17-0.20)	0.19 (0.17-0.21)	6 (4.4-7.6)
RW	0.06 (0.04-0.09)	0.12 (0.07-0.19)	0.16 (0.15-0.17)	0.13 (0.12-0.15)	32.4 (23.8-47.1)
RA	0.09 (0.06-0.12)	-	-	-	13.7 (10.3-18.4)
RB	0.08 (0.06-0.10)	0.13 (0.08-0.20)	0.25 (0.23-0.26)	0.23 (0.22-0.25)	55.3 (37.3-97.3)
LX	0.19 (0.12-0.30)	0.56 (0.31-0.77)	0.53 (0.49-0.57)	0.45 (0.40-0.50)	1.2 (1-1.5)
EM	0.42 (0.29-0.56)	0.46 (0.17-0.72)	0.72 (0.67-0.76)	0.65 (0.59-0.71)	0.7 (0.5-1)
SA	0.68 (0.51-0.83)	0.48 (0.20-0.74)	0.67 (0.62-0.72)	0.70 (0.64-0.75)	-1.6 (Inf-Inf)
TD	0.03 (0.00-0.13)	0.46 (0.23-0.68)	0.62 (0.57-0.67)	0.58 (0.52-0.63)	-1 (Inf-Inf)
Mean	0.11	0.22	0.29	0.27	90.78

[1] Negative estimates are indicative of insufficient sample size, with the size correction being greater than the r^2 value calculated for the data.

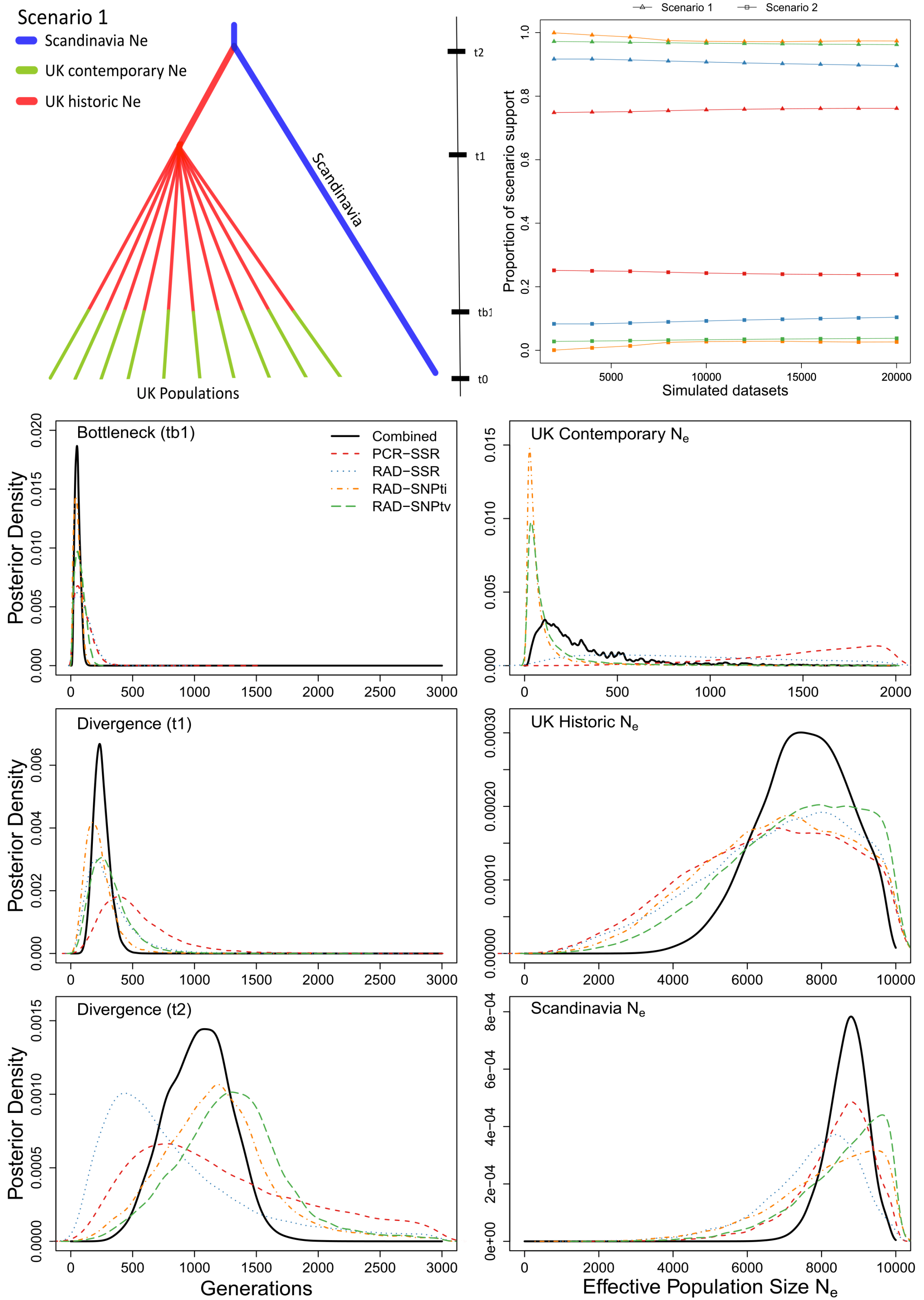


Figure 3.4 Posterior demographic density distributions for all marker sets independently and with information combined (black line). Line graph shows higher support for Scenario 1.

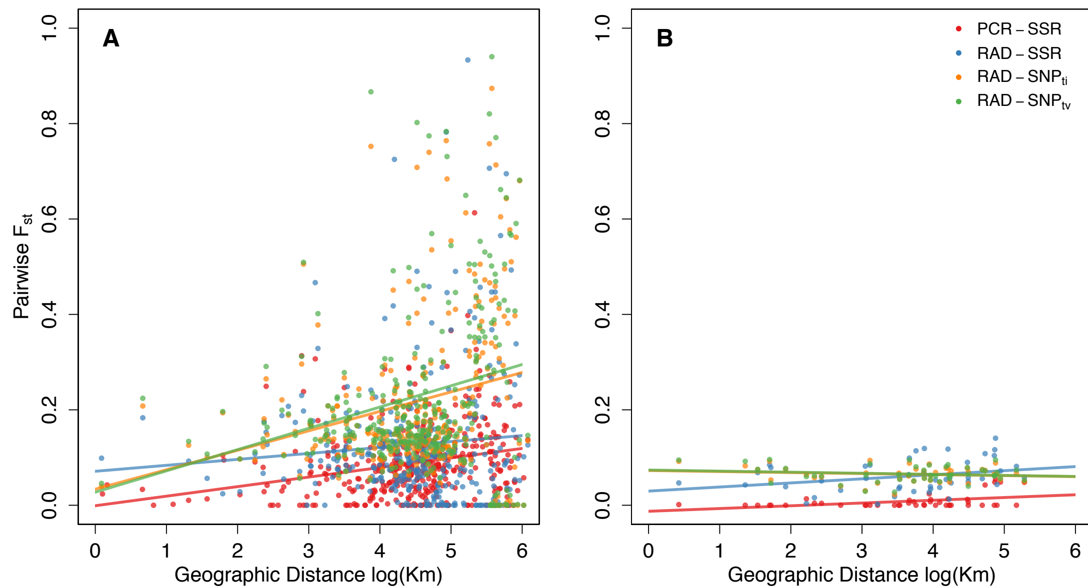


Figure 3.5 Scatter plot of linearized pairwise F_{ST} versus pairwise log geographic distance for all study populations for A) UK and B) Scandinavia. Weak isolation by distance ($p<0.05$) was detected in UK and Scandinavian populations using PCR-SSRs. All other relationships were non-significant.

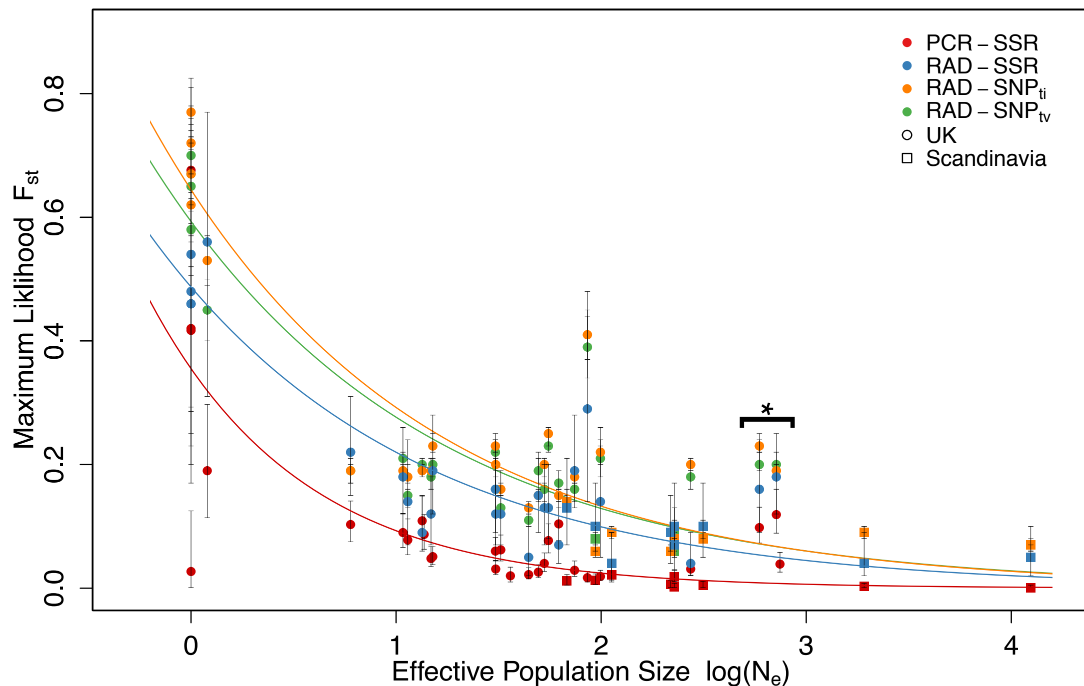


Figure 3.6 Scatter plot of the relationship between log effective population size (N_e) and maximum likelihood F_{ST}, for each marker type, using all datasets across both regions. Regression shows increased differentiation in smaller populations.Populations MU1 and MU2, denoted by '*' are outliers with a large effective population size, but also moderate differentiation.

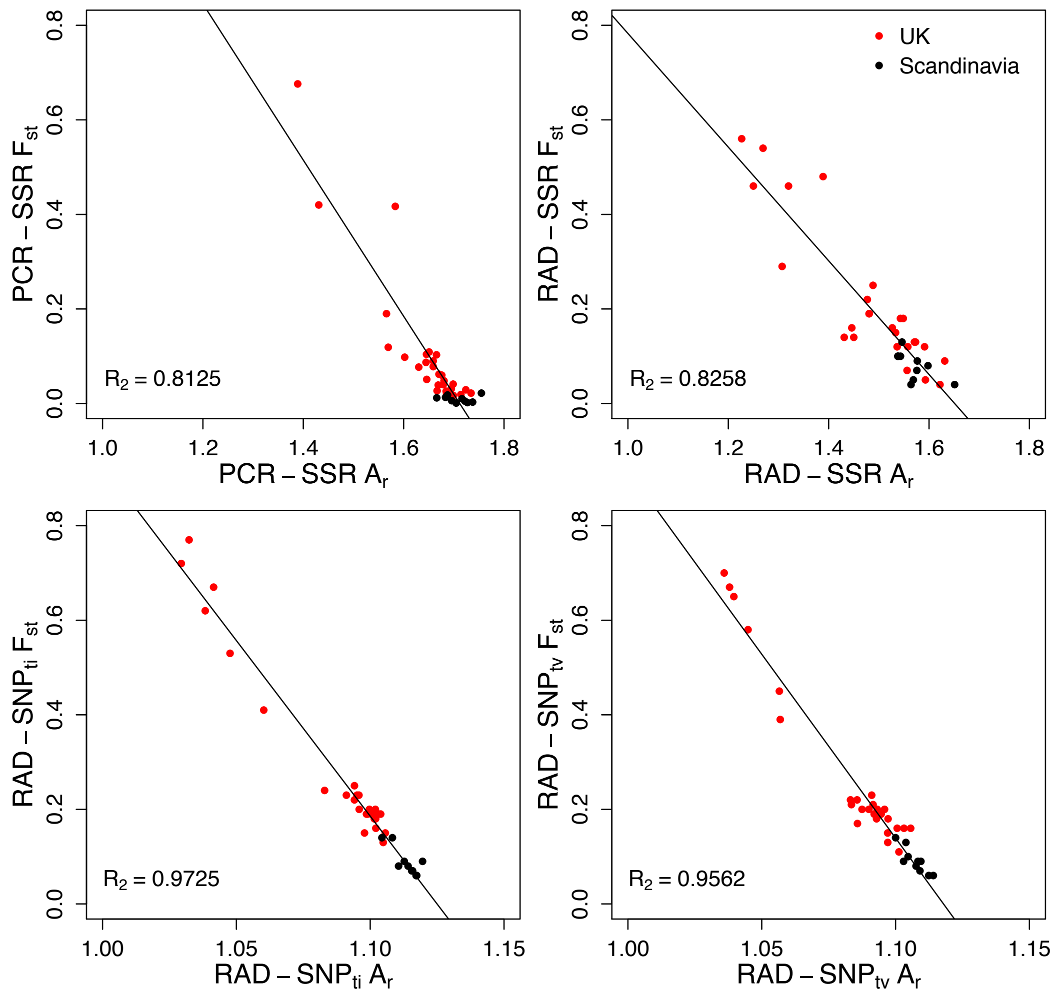


Figure 3.7 Regression of rarefied allelic richness (rarefied to equivalent of one diploid individual), against maximum likelihood F_{ST}, showing that highly differentiated populations also have low genetic diversity.

3.4 Discussion

We generated population genomic datasets for dwarf birch populations in the UK and Scandinavia using four types of markers: 18 PCR-SSR loci with a total of 421 alleles, 34 RAD-SSR loci with 354 alleles and biallelic RAD-SNP markers with a core dataset of >8000 loci, and a transition to transversion ratio of 11:10. We successfully genotyped 1066 individuals with the PCR-SSRs and 187 individuals with the RAD-based markers. The number of loci, alleles and samples for these marker sets reflect their abundance in the genome, the dynamics of their mutation and the ease of assaying them.

3.4.1 Agreement among marker sets

In broad terms, the four marker datasets gave very similar results. All bar two comparisons identified significantly higher allelic richness, private allele frequency, observed and expected heterozygosity in Scandinavia than the UK (Figure 3.2, Table 3.5). They all showed greater population differentiation in the UK than in Scandinavia and all markers across almost all populations displayed an observed heterozygosity deficit, suggesting they are not at equilibrium (Table 3.5). Within the UK, all markers identified highly significant trends with latitude in both expected heterozygosity and inbreeding (Figure 3.2). In addition, they displayed rapidly accelerating maximum likelihood F_{ST} values at lower population sizes (Figure 3.6), suggesting that genetic drift is predominant in the evolution of very small populations of dwarf birch in the UK. In ABC modeling of population history (Figure 3.4), the four marker datasets independently supported the same scenario, with combined posterior estimates across markers improving inference of demographic events and confidence in the conclusion. Of the six estimated parameters, the four marker datasets conflicted only in the estimate of contemporary UK N_e.

3.4.2 Complimentary differences among datasets

The markers also proved to be complementary in a manner that reflects their differences in mutation rates. Here, we expect mutation rates to display the following relationship: PCR-SSR > RAD-SSR > RAD-SNP$_{ti}$ > RAD-SNP$_{tv}$, and this is reflected in our data. We observed a higher number of alleles per locus for PCR-SSRs than RAD-SSRs, likely due to their longer repeat sequences being more mutable (repeat length of RAD-SSRs was limited by the total length of the raw Illumina reads minus adaptors, ~90 base pairs). We also found the number of transition vs transversion SNPs was similar post-filtering, whilst very low frequency transitions were more abundant in the unfiltered dataset consistent with their higher propensity to mutate. These differing mutation rates can be used to explain the observed patterns in population summary statistics (Tables 3.4 and 3.5).

In the context of increased fragmentation, our expectation was that UK populations should display higher F_{ST} suggesting that they are isolated with low migration between populations. Across markers, we found higher F_{ST} in the UK, but little evidence of significant isolation by distance in pairwise F_{ST} values, except in PCR-SSRs where a higher value of Mantel's *r* in Scandinavia has recently been shown to provide evidence of a higher local migration rate (Meirmans 2015). Importantly, our approach utilizing multiple markers allows us to distinguish migration from mutation, which may sometimes confound signatures of genetic differentiation. Migration (gene flow) between populations should have the effect of reducing differentiation across all marker datasets, whereas homoplasy due to recurrent mutation should only be apparent at loci with high mutation rates. Thus in Scandinavian populations where the difference in F_{ST} estimates between marker datasets is smaller, we can infer that migration is moderating differentiation. Whereas in the UK, differences among datasets in F_{ST} estimates are greater, especially towards the Southern range edge where populations are highly separated, hence gene flow diminished.

As expected, F_{ST} estimates at loci with higher mutation rates have lower F_{ST} values (Coates et al. 2009). This effect is diminished in our RAD-SSRs perhaps because of lower mutation rate, but also because they are less prone to apparent homoplasy because, unlike PCR-SSRs where alleles are distinguished in the basis of amplicon length, we have the sequence reads for RAD-SSRs and can therefore distinguish alleles that differ in sequence but are the same length.

We expected all markers to give the same F_{IS} estimates, as nonrandom mating should affect all markers equally, but unexpectedly PCR-SSRs gave much lower F_{IS} values than other markers (Tables 2 and S6). This effect may be attributable to an ascertainment bias, since the RADseq datasets contain low frequency alleles, which, being present in single copies cannot be present as homozygotes, thus depressing standard F_{IS} estimators. Alternatively, we raise the possibility of RAD allele drop out (Gautier et al. 2013) resulting in an underestimation of heterozygosity. It is possible that longer PCR-SSR loci, where primer binding sites are further from the variable motif or polymorphism (total length 3-400bp), are less prone to these. Thus the PCR-SSRs yield a lower estimate of F_{IS} than the other three marker sets based on shorter ~100bp sequences, despite having different marker types.

We also expected all markers to give a similar discrimination between clusters in Bayesian structure analysis, separating UK and Scandinavian populations, with the possible exception of PCR-SSRs, given that these are more prone to homoplasy. However, we found RAD-SSRs to perform very poorly in this discrimination (Figure 3.3). The greater success of PCR-SSRs may have been due to a larger sample size than for the RAD-SSRs. Or the RAD-SSRs may have performed poorly due to higher levels of missing data. Another possibility is that homoplasy is actually more common in our RAD-SSRs, given that they are based on shorter repeat lengths in which reverse mutations might be more common.

Of the six estimated parameters in our ABC analysis, the four marker datasets conflicted only in the estimate of contemporary UK N_e, with a smaller value estimated based on the RAD-SNP markers. We attribute this to ascertainment bias in the selection of SSR loci, which over-

represents more highly variable loci and thus inflates the estimate of contemporary N_e. This is an issue that could be overcome with improved unbiased sampling of RAD-SSRs.

3.4.3 Present and past status of dwarf birch in the UK

Our data clearly show that UK dwarf birch is genetically differentiated from Scandinavian dwarf birch, and currently exists in small populations that are genetically differentiated from one another. Our ABC models give estimates of generation time rather than absolute dates. It is difficult to estimate the generation time of *B. nana* as individuals can set seed in as little as 18 months in a glasshouse (personal observation), whilst successful propagation in the field appears to depend largely on prevailing conditions and is highly variable across years and sites. If we assume that the end of the last ice age was around 10800 years ago, and this is when the first divergence occurred between the ancestral UK and Scandinavian dwarf birch populations, this would fit the estimate of our model if the generation time is 10 years, which is biologically plausible. Thus, at the end of the last ice age, we imply colonization with a large effective population size. By 2000 years ago, consistent with our estimate of *t1*=233 generations, the literature suggests large scale anthropogenic activity including lowland agriculture had begun to shape the distribution of woody plants in Scotland (Smout, T. et al. 2005; Godwin 1975; Tipping 1994). The isolation of highland dwarf birch habitat by lowland agriculture and forest clearance may have permitted a limited amount of differentiation to begin to accrue in separate populations, but based on ABC analysis and previous studies of widespread introgression in downy birch from dwarf birch (Wang et al. 2014), it is likely that a large overall effective population size still persisted. Most recently, accelerated forest clearance in the 17th and 18th century, combined with a shift to livestock farming during the highland clearances (Dodgshon & Olsson 2006) and an increase in wild herbivores, partly as a result of the removal of wild predators (Holl & Smith 2007), could plausibly have impacted dwarf birch numbers. It is likely that one or several of these factors contributed to the recent severe bottleneck in dwarf birch estimated at *tb*=50 generations and the small effective population size that we

observe today. By comparison, in Scandinavia, the topology is substantially flatter and grazing pressure is lower, so a lack of both anthropogenic activity and habitat isolation separated by unsuitable areas contrastingly led to the low levels of genetic differentiation and large effective population size observed in the Scandinavian lineage today.

3.4.4 Conservation implications for dwarf birch

Together, these findings have two important implications for the proactive conservation of dwarf birch in the UK. First, our results give a clear picture of severe population decline and fragmentation in the UK, relative to larger, healthier populations in Scandinavia. Analysis of genetic differentiation shows genetic drift to be the dominant evolutionary process driving contemporary differentiation in many UK *B. nana* populations, as a result of small effective population size. This interpretation is supported by field observations, linkage disequilibrium and demographic estimates of population size, which all point to *B. nana* populations in the UK being extremely small (~1-2 orders of magnitude smaller than Scandinavia). Furthermore, the significant negative relationship between rarefied allelic richness and genetic differentiation (Figure 3.7) suggests a reduction in diversity in differentiated populations and that the degree of divergence depends on the strength of genetic drift. Thus we can expect rapid genetic deterioration of small UK populations. Indeed in the context of climate change, several authors have highlighted the importance of receding edge diversity (Hampe & Petit 2005; Provan & Maggs 2012), and we also note the high relative frequency of private alleles in the southernmost populations, based on PCR-SSR and RAD-SSR markers, suggesting that there may be important diversity that will almost certainly be lost if these populations cannot be conserved.

Second, in the medium term population subdivision and differentiation has actually helped maintain genetic diversity in the UK, despite substantial reductions in regional effective population size. The high levels of F_{ST} identified in this study indicate that genetic diversity is partitioned over populations, rather than within populations. In such situations genetic

diversity is relatively resilient. For example if alternative alleles are fixed in different populations they will persist. However in the context of future anthropogenic pressures and climate change, genetic diversity is only useful for adaptation where there is the opportunity for gene flow to allow beneficial alleles to spread. This effect is perhaps more likely in tree species on account of longevity and potential for occasional long distance gene flow (Kremer et al. 2012). This perspective give cause for optimism by raising the possibility that future adaptability of this species would be enhanced by artificially restoring gene flow between populations, perhaps by culturing seedlings within, or transplanting among populations.

Although genetic variation is highly differentiated among UK populations, levels of genetic diversity in some populations at the center of the range are comparable to populations surveyed in Scandinavia. Similarly, the restoration of gene flow among populations in the UK could produce levels of diversity approaching those in Scandinavia. We found evidence of severe genetic drift only at remarkably low population sizes (Figure 3.6); hence, even very modest interventions may be effective.

An unexpected finding of our study is that in all marker sets inbreeding increased with latitude in the UK, displaying lower values in smaller more fragmented populations. Dwarf birch naturally forms a dense montane scrub habitat, thus a possible explanation is severely spatially limited gene flow in this species meaning reproduction is frequently with close neighbours. Where effective population size has declined this may result in less dense populations (consistent with field observations), thus inbreeding may decline as populations reduce.

3.4.5 Broader utility of our approach

Whilst, the potential for orders of magnitude more data from SNP analyses based on modern NGS technology is widely appreciated, comparative studies combining different categories of population genetic markers remain sparse (examples include Bradbury et al. 2015; Jeffries et al. 2015). Here we have found that for studies focusing on neutral differentiation and drift, a

modest number of SSR markers produce results that are comparable with thousands of SNPs (see also Hodel et al. 2016). We suggest that RAD-SSRs could be more widely used as they overcome many of the limitations of PCR-SSRs, yielding a high mutation-rate marker set, with a well-specified ascertainment process that does not entail additional sequencing. In this way, we were able to exploit markers with different mutation rates to explore different demographic scenarios. Comparing different types of markers can alert the researcher to biases introduced by the different ascertain processes for different markers (in our case possible effects on demographic estimates of contemporary population size and inbreeding). Particularly for conservationists where management interventions based on population data may strongly influence the fate of species, robust conclusions based on multiple independent data sets are less likely to be subject to unidentified technical or analytical artifacts.

3.4.6 Concluding remarks

Our findings generally support the view of Lowe et al. (2015) suggesting a level of resilience in tree populations to moderate fragmentation. Considering the demographic decline and isolation of UK dwarf birch populations, the maintenance of modest levels of diversity in the center of the range is encouraging. Whilst our results provide improved insights into the consequences fragmentation, they underscore the need for management intervention to mitigate the local loss of genetic diversity.

Chapter 4: Genomic assessment of adaptive potential in dwarf birch under present and future climates

Summary

Variation in patterns of natural selection across a species' range can lead to local adaptation and the genetic differentiation of populations. The degree of local adaptation can be considered as a measure of concordance between an organism's genotype under selection and the environmental conditions to which it is exposed, with adapted populations displaying a fitness advantage under local conditions. Where populations have become fragmented with reduced effective population size, gene flow and genetic diversity, the potential for local adaptation to evolve is predicted to be weak. However adaptation across environmentally heterogeneous distributions is likely to be important for species persistence, particularly under scenarios of future climate change. Here, we propose a method to measure the degree of local adaptation across natural populations of the montane plant dwarf birch (*Betula nana*) by estimating allele frequency deviation from a theoretical genotype-environment optimum. We find substantial variation across the species range, and show that deficit in adaptive potential (DAP) for annual mean temperature is significantly negatively correlated with catkin production. In an assessment of the relative importance of environmental variables driving local adaptation we find that both species distribution modeling and an analysis of genotype-environment correlations report a conserved set of environmental variables. Finally, by measuring the potential for assisted gene flow to minimize the adaptive deficit of dwarf birch population to local environmental conditions, we conclude that the optimum distribution of adaptive genetic diversity for restoration, if desirable, could require composite provenancing of genotypes from across the species' UK range.

4.1 Introduction

The rate of global environmental change is increasing (Smith et al. 2015). Species adapted to current local environments may respond to changes and avoid extinction by: i) migrating to track favorable conditions (Aitken et al. 2008; Meier et al. 2012), ii) exhibiting phenotypic plasticity to tolerate a range of conditions (Gratani 2014; Nicotra et al. 2010), or iii) evolving adaptations from standing genetic variation, new mutations, or gene flow to maintain fitness (Aitken et al. 2008; Alberto et al. 2013).

There is abundant recent evidence that species have survived considerable climatic changes in the past by migration; for example, during the last glaciation European tree species distributions shifted south, tracking more favourable conditions (Bennett et al. 1991; Willis et al. 2000; Birks & Willis 2008). However, numerous studies have cast doubt on the potential for migration to accommodate species response to contemporary anthropogenic climate change, particularly in plants (Hampe & Petit 2005; Zhu et al. 2012; Corlett & Westcott 2013). In some cases species may lack the dispersal ability to keep pace with accelerated rate of climate shifts (Svenning 2003; Loarie et al. 2009; Meier et al. 2012), there may be an absence of potential habitat at higher latitudes (McKenney et al. 2007), and altitudes (Engler et al. 2011) or suitable new habitats may be separated by unsuitable intervening spaces (Meier et al. 2012). Thus an essential starting point in predicting the effects of environmental change on species distributions is understanding the degree to which populations are currently locally adapted (Savolainen et al. 2007; Savolainen et al. 2013; Blanquart et al. 2013; Harrisson et al. 2014).

Local adaptation is most reliably detected through reciprocal transplant experiments, which can reveal genotypic effects on fitness traits (Kawecki & Ebert 2004; Leimu & Fischer 2008; Gibson et al. 2016) and subsequent functional validation of candidate genes in controlling biological processes (Eckert et al. 2009; Ingvarsson et al. 2016). However, this approach is often not feasible for rapid assessment of wild non-model plant populations with long

generation times or urgent conservation of adaptive genetic resources (Volkmann et al. 2014). A recently developed alternative is environmental association analysis (EAA) (reviewed in Rellstab et al. 2015), which has become an increasingly powerful tool in landscape genomics due to the resolution afforded by reduced representation sequencing (RADseq) (Andrews et al. 2016). EAA seeks to identify genetic differentiation resulting from biotic or abiotic selection, by correlating allele frequencies with environmental variables whilst accounting for confounding neutral genetic structure (Savolainen et al. 2013; Günther & Coop 2013; Rellstab et al. 2015). This is based on an assumption that similar environmental selective pressures result in similar genomic patterns arising either through gene flow, parallel adaptation, or selection on ancestral standing variation (Jones et al. 2012; Renaut et al. 2014). Thus a replicated signature of selection across many independent populations provides strong evidence for an adaptive role in candidate loci (Renaut et al. 2014). Whilst this approach may identify correlation between environment and genotype, it cannot prove causation, thus this approach may be considered complimentary to longer-term reciprocal transplant experiments.

Several recent studies have reported evidence of local adaptation in a variety of tree species (Keller et al. 2012; Chen et al. 2012; Dillon et al. 2014; Geraldes et al. 2014; Suarez-Gonzalez et al. 2016; Wang et al. 2016) with the degree of adaptation across wild populations likely to be influenced by a number of factors (Siol et al. 2010). Specifically, local adaptation is expected to be more common in populations with moderate gene flow, hence avoiding genetic swamping at high gene flow (Polechová & Barton 2015; though see Kremer et al. 2012) whilst not impeding the spread of adaptive alleles at low gene flow. Second, local adaptation is expected to be weaker in populations with a small effective size because genetic drift results in the random fixation of alleles irrespective of weak selection pressures (Ellegren 2009).

Whilst local adaptation may arise within a species through the mechanisms outlines above, alternatively it is also possible that introgression through interspecific hybridization may contribute adaptive genetic material (Arnold & Martin 2009). Though the degree to which this

occurs in the wild is largely unknown (Brennan et al. 2014). It has been hypothesized that interspecific gene flow may increase species' ability to respond to climate change (Hamilton & Miller 2015) and that this may represent a neglected management opportunity for the conservation of evolutionary potential (though see Gomez et al. 2015). In European *Betula*, a latitudinal cline of introgression has previously been observed, with interspecific hybridization more common at the advancing range edge (Palme et al. 2004a; Maliouchenko et al. 2007; Wang et al. 2014; Eidesen et al. 2015). It has been suggested that transfer of adaptive alleles may facilitate the invasion of new species, or improve tolerance to changing climate (Taylor et al. 2015; Suarez-Gonzalez et al. 2016).

Over time, population allele frequencies and resulting phenotypes respond to natural selection to remediate the discordance between adaptive variation and prevailing selection pressures. However, where long lived species, with small populations and limited potential for migration are exposed to rapid environmental change, adaptive potential may be limited (Alberto et al. 2013), with genetic drift or extinction occurring before adaptation can take place. A management alternative is assisted gene flow, here defined as the managed movement of individuals or gametes between populations (Aitken & Whitlock 2013). Assisted gene flow theoretically permits the matching of optimal genotypes from elsewhere in the species' distribution to specific local environmental conditions. Where appropriate, this has the potential to inform sourcing of seed stock for reforestation programs (Boshier et al. 2015) and mitigate maladaptation to future climate (Havens et al. 2015; Gibson et al. 2016; Aitken & Bemmels 2016).

Two key requirements must be met however for effective and appropriate assisted gene flow. The target species should be distributed across heterogeneous habitats to carry diverse genetic material, and assure that one extremity of a species distribution is already exposed to environmental conditions that other populations may be exposed to in the future, thus providing an indication of adaptive potential (Olson et al. 2013). Second, it is important to

accurately predict future species distributions and identify major selection pressures. This particular requirement can be achieved with the application of species distribution models (SDMs) (Maguire et al. 2015). SDMs infer relationships between species and their environment, permitting the projection of species ranges under future climate scenarios (Elith & Leathwick 2009) and the characterization of explanatory environmental variables (Rolland et al. 2015; Fitzpatrick & Keller 2015; T. Wang et al. 2016) (see Chapter 2). Thus far, few studies have attempted to incorporate patterns of local adaptation with the broad predictive capacity of SDMs (though see Rolland et al. 2015), or compared the relative importance of environmental variables identified across the two methods. Indeed, current models assume species to be genetically homogenous and ecologically uniform in tolerance to climate across their range (Hällfors et al. 2015; Alberto et al. 2013). Thus synthesis of population genetic and SDM outputs may lead to enhanced prediction of species' response to environmental change.

In this study we combine SDM and EAA based on thousands of genome wide SNP markers, together with phenotypic fitness data, to predict the adaptive potential of the dwarf birch, a nationally scarce montane tree that has experienced an accelerated decline across the UK in recent decades (Aston 1984; Wang et al. 2014). Dwarf birch, like many tree species, is the focus of a conservation program to restore populations, delimit management units and prioritise important genetic diversity (Koskela et al. 2013). To guide this strategy, we (i) utilized the SDM from chapter 2 to projected the species' range under present and future climate scenarios and characterized environmental factors that determine its distribution, validating this with phenotypic metrics of reproductive output; ii) we empirically test for concordance of genomic markers with environmental variables using SDM and EAA analyses, and attempt to validate candidate adaptive loci with blast and RNA expression analysis; iii) we screen for evidence of adaptive introgression from *B. pubescens;* and (iv) we test how the deficit of adaptive potential compares to established methods to rank the conservation value of

populations. Finally, (v) we use the combined information to design a hypothetical model of assisted gene flow to maximize adaptive genetic diversity and therefore adaptive potential.

4.2 Materials and Methods

In this chapter we utilize the *B. nana* species distribution model developed in Chapter 2, together with associated percentage contribution values of explanatory environmental variables. Similarly, the genetic samples used in this study are a UK subset of those described in Chapter 3. Briefly, DNA was extracted from 130 *B. nana* individuals from UK populations (see Table 3.3) and submitted to Floragenex (Oregon, USA) for 100bp single end RAD sequencing using the enzyme PstI (see Chapter 3 for RAD sequencing protocol). To test for evidence of introgression from *B. pubescens* into *B. nana* genomes, we also collated 138 *B. pubescens* samples from 45 locations across the UK (Table 4.1). These were reported in Zohren et al. (2016) and sequenced using the same methodology at the GenePool genomics facility (University of Edinburgh, UK) to yield 50bp single end reads. We used two separate approaches to align, call and filter SNPs for local adaptation and introgression analyses respectively:

First, to assess local adaptation across *B. nana* populations, raw sequencing reads were filtered using the *process_radtags* module of Stacks v1.35 (Catchen et al. 2013) and then aligned to the V4 *B. nana* genome retaining only reads that align uniquely (Wang et al. 2013) using Bowtie2 (Langmead & Salzberg 2012) and the Stacks *ref_map.p*l pipeline. SNPs were called with a minimum depth (-m) of 5, a bounded model (0.001-0.01 error rate) and default values for other parameters. Corrections were made using the *rxstacks* pipelines and SNPs recalled with a minimum log likelihood of -20. Using the *populations* module, we filtered for loci present in ≥30% of *B. nana* individuals in ≥8 populations, and a minor allele frequency >0.05. Here this dataset is termed *SNP_env*. Second, to assess evidence of introgression from *B. pubescens* into

B. nana we aligned loci from both species to a consensus *Betula* reference genome derived from 12 species' RAD loci and their flanking regions in the v4 *B. nana* genome in the software CLC bio (Qiagen Aarhus, Denmark). This approach aimed to improve our ability to distinguish true introgression from homologs at loci present in both study species.

Table 4.1 *Betula pubescens* sampling locations and number of individuals

Population	Latitude	Longitude	Samples
Berriedale Wood, Orkney	58.89	-3.38	3
Betty Hill	58.53	-4.21	2
Loch Linnhe	58.49	-4.66	1
The Crawford Population	58.48	-4.22	2
Castle Varich Woods	58.47	-4.42	4
Tongue	58.46	-4.42	4
Ben Loyal	58.42	-4.42	22
Lairg	58.03	-4.44	3
Drumrunie	57.99	-5.11	5
Ardgay	57.88	-4.36	4
Ben Wyvis	57.69	-4.63	2
Cromarty	57.68	-4.01	3
Urqhart Castle	57.32	-4.45	1
Lynemore	57.28	-3.55	2
Aviemore	57.12	-3.90	4
Gairnshiel Lodge	57.10	-3.15	1
Kingussie, Highland	57.08	-3.98	3
Mar Estate ne Braemar	57.01	-3.54	3
Glen Muick	57.00	-3.08	2
Loch Muick	56.92	-3.21	6
Ben Gulabin (nw)	56.83	-3.49	3
Crianlarich	56.42	-4.51	1
Bankhead Moss	56.28	-2.90	1
Loch Lomond	56.23	-4.70	1
Johnstonebridge, Dumfries	55.22	-3.42	4
RoseberryTopping	54.51	-1.11	4
BirkPark, Yorks	54.39	-2.00	3
RichmondQuarry, Yorks	54.39	-1.83	3
Flaxby, North Yorkshire	54.01	-1.39	2
BoltonAbbey	54.00	-1.89	3
ScoutCamp, Lancs	53.80	-2.41	3
GlossopWood, Derby	53.43	-1.95	3
Woodbastwick Marshes	52.68	1.46	3
S of Attleborough	52.49	0.99	2
Eccles Car, Norfolk	52.45	1.00	3
RytonWood, Warwick	52.35	-1.45	3
BreconBeacons2	51.85	-3.18	3
Danbury	51.71	0.57	3
Capel, Kent	51.17	0.34	2
Long Copse	51.11	-0.93	1
Williand Wood	51.04	-0.61	1
Cootham	50.92	-0.48	2
N of Beaulieu	50.85	-1.45	2
DartmoorBog	50.52	-3.82	3
DartmoorBurrator	50.49	-4.05	3

	Total	139

B. pubescens presents additional challenges to accurate genotyping as it is polyploid, which frequently excludes species from population genetic analysis due to difficulties in measuring allele dosage (Dufresne et al. 2014). Thus SNP calling was performed using a custom script published in Zohren et al. (2016) that inputs read counts and base qualities together with ploidy, in this case tetraploid, to obtain maximum likelihood estimates of allelic ratios in polyploidy organisms. Subsequently we constructed a mixed ploidy dataset using filtering constraints to permit identification of low frequency alleles that may be recently introgressed into *B. nana* or present in only a few populations. Retained *B. pubescens* SNPs were present in >50% of individuals, with a minor allele frequency of >0.05, whilst for *B. nana* SNPs there was no minor allele frequency filter and loci had to be present in five or more populations. For downstream Bayesian clustering analysis, we coded all individuals as tetraploids, using two lines of missing data for *B. nana* individuals. Here this dataset is termed *SNP_int*.

4.2.1 Genomic signatures of local adaptation

To detect loci putatively under diversifying or balancing selection, we first use BayeScan (Foll & Gaggiotti 2008) to compare allele frequency differences among populations. BayeScan assess evidence that loci are F_{ST} outliers based on comparison of the posterior probability of neutral and selection models incorporating population-level F_{ST} and locus-specific F_{ST} respectively. Narum and Hess (2011) found that of several F_{ST} outlier tests, BayeScan displayed amongst the lowest type I and type II errors. Our analysis was performed with 50,000 iterations thinned every 10, with 20 pilot runs and a burn-in of 50,000 iterations.

Whilst F_{ST} outliers identified in BayeScan are candidate loci of adaptation, processes other than local adaption can lead to F_{ST} outliers, such as selection due to deleterious alleles, selective sweeps, hybrid zones, genetic drift and stochastic processes in expanding populations (Bierne

et al. 2013). Thus, we combine outlier detection methods with environmental association analysis (EAA) implemented in Bayenv2 (Günther & Coop 2013). Bayenv2 incorporates neutral genetic structure using a covariance matrix based on neutral markers and attempts to identify correlations between outliers and environmental gradients, potentially reducing false positives (De Mita et al. 2013). We first computed a null covariance matrix of relatedness between populations in Bayenv2, over 100,000 iterations and five independent runs, excluding loci detected as under putative selection in BayeScan. We then tested all loci under an alternative model where allele frequencies are determined by a combination of the covariance matrix and an environmental variable. The posterior probability was assessed using Bayes factors (BF), with log10 posterior odds ratio values >1 defined as strong support by Jeffreys' scale of evidence (Jeffrey 1961). To validate our findings we averaged BFs over five independent runs as recommended by Blair et al. (2014), and following Günther & Coop (2013) we retained loci as good candidates if in addition to a high BF, they also fell in the top 5% of Spearman correlation coefficient (p) values. We performed this analysis twice, the second time excluding seven individuals that were identified as subject to severe genetic drift. Here selection is unlikely to be a major contributor to allele frequencies, thus this was considered to be a potentially more conservative analysis.

4.2.2 Introgression

To test for evidence of adaptive introgression from *B. pubescens* into *B. nana* populations, we employed three approaches. First, we used the Bayesian clustering algorithm implemented in the software STRUCTURE v2.3.4 (Pritchard et al. 2000). All individuals were coded as tetraploid, to permit combined analysis of 265 individuals in the *SNP_int* mixed ploidy dataset. Parameters consisted of the admixture model, correlated allele frequencies, 100,000 burn-in iterations and 100,000 repeats. We assumed k=2 and conducted 20 independent runs. Second, we filtered for alleles that were variable, but at >90% frequency in *B. pubescens*, whilst being at a low frequency in *B. nana*, (that is, fixed for the alternative allele in most populations).

Where a *B. pubescens* allele occurs in a *B. nana* individual, this suggests evidence of putative introgression or a novel mutation. Finally, we test for evidence that introgression displays spatial structure. Contrasting levels of introgression in different populations would provide further support for detection of real introgression rather than shared ancestry or novel mutations. To assess evidence for adaptive role in introgressed loci, we compared putative introgressed loci with those identified in outlier tests and environmental association analyses.

4.2.3 Relationship between genomic and environmental data

We investigated relationships between geographic distance, environmental distance (see chapter 2) and genetic distance for adaptive and neutral loci using mantel and partial mantel tests implemented in the package Vegan (Dixon 2003) in R software. Pairwise F_{ST} was computed in Arlequin v3.5.2 (Lischer 2010) for the whole *SNP_env* dataset, and also separately for neutral and putative adaptive loci. Pairwise geographic and environmental distance was calculated using Euclidian distance in the package *fields.* We conducted the following tests: Geographic distance and environmental distance, geographic distance and F_{ST} and finally environmental distance and F_{ST}. In each case, we tested neutral and adaptive loci separately. We also performed a partial mantel test between environmental distance and F_{ST}, controlling for geographic distance. Finally, we conducted linear regression on logit transformed percentage contribution of environmental variables against the square root transformed number of SNPs associated with each environmental variable, to test the level of concordance between the two methods.

4.2.4 Gene ontology and gene expression

To assess putative function in genes closely linked to SNP outliers identified using Bayenv2, we extracted up to 10,000bp flanking either side of the candidate loci from the *B. nana* reference genome. Where contig or scaffold length was <20,000bp, we retained the whole contig sequence. We conducted *Blastx* searches against the NCBI *nr* public database and gene ontology (GO) annotation in the Blast2Go pipeline (Conesa, A. & Gotz 2007), using an e-value

threshold of $1x10^{-10}$. To determine whether molecular functions, biological processes or cellular components of markers were disproportionately represented we compared candidates to a randomly sampled dataset of non-outlier loci regions. Enrichment analysis was performed using a two-tailed Fisher's exact test, with a false discovery rate of 0.05. We performed an identical analysis for loci identified as introgressed from the *SNP_int* dataset, with a Fisher exact test against loci present in both species, but without evidence of introgression.

Finally, to provide an additional line of evidence for whether candidate loci are linked to genes, we compared loci to a database of RNA expression sequences derived from *B. nana* tissues. Briefly, RNA was extracted from fresh leaves and flowers using a modified Qiagen RNAeasy Plant Mini Kit, incorporating additional CTAB and Phenol-Chloroform steps. Sequencing was performed at the Genome Centre of Barts and the London School of Medicine and Dentistry, to generate 100bp paired-end reads with an average insert size of 280bp and mapped to the *B. nana* reference genome using Trinity software (Manfred et al. 2013). Here we test for significant overrepresentation of RNA alignments among candidate adaptive loci, compared to a random sample of sequences.

4.2.5 Deficit in adaptive potential

We present a metric for the degree of genotype-environment concordance, which we tentatively term "Deficit of Adaptive Potential" (DAP). Retaining candidate environmentally-associated loci, we performed best linear unbiased predictions (BLUP) on raw allele frequencies and plot the product of estimated effect size and allele frequency against the corresponding environmental variable, using the R package rrBLUP (Endelman 2011). We restricted this analysis to the nine environmental variables with >6 associated loci and hypothesized that the resulting linear regression line represents the optimal allele frequency across the measured range of environmental values. We make the assumption that relative population fitness is to some degree determined by the frequency of putative adaptive alleles. For each population we calculate DAP as the mean of the absolute residuals along the

regression line for each environmental association. Where an allele was significantly associated with more than one environmental variable, we included only the association with the largest BF value.

To assess how DAP is likely to change in the future, we used the slope of each regression model to predict future optimum allele frequencies based on the projected environmental values of each climate change scenario for each study population (see Figure 4.1). This is an approach conceptually similar to that described for phenotypic traits in St Clair & Howe (2007) and recently termed RONA (risk of non-adaptedness) in Rellstab et al. (2016). The new DAP_{future} value estimates the adaptive allele frequency change necessary for the population to be optimally suited to future climate based on the current model, though we note assumptions and potential limitations in section 4.4.4.

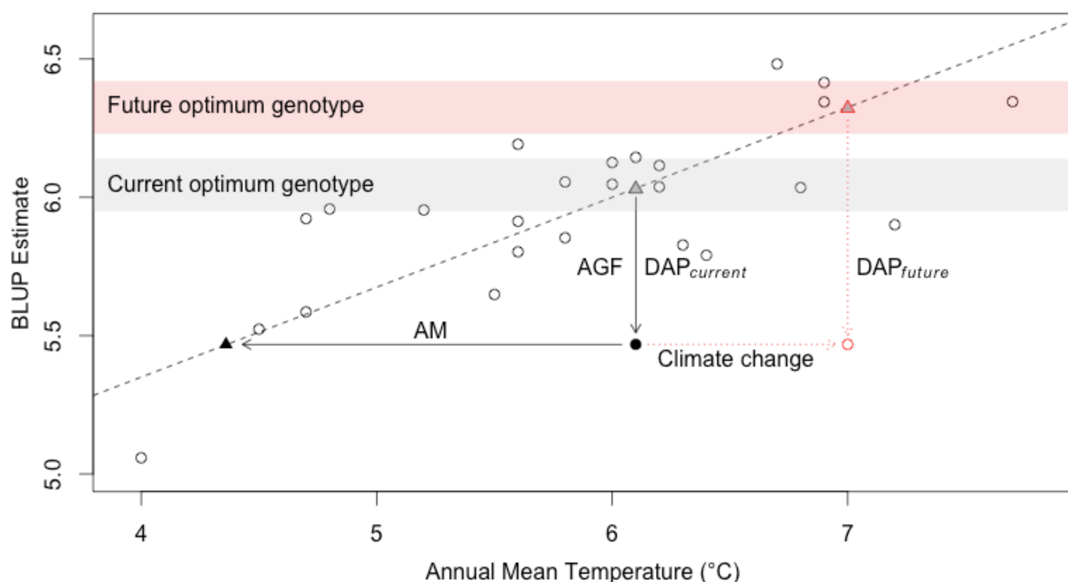


Figure 4.1 Schematic diagram of Deficit in Adaptive Potential (DAP) under current and future climate, presented on a genotype-environment BLUP plot for Annual Mean Temperature with dashed line denoting theoretical optimum. AGF shows potential assisted gene flow from a donor population with a current optimum genotype. Alternatively, assisted migration (AM) could take the form of translocation of individuals from the focal population to a location with environmental conditions matching their genotype. Horizontal red line indicates the magnitude change in Annual Mean Temperature likely to be experienced by this population, with vertical red line DAP_{future} denoting the increased allele frequency deficit under this

scenario. Grey and red bands indicate suitable candidate donor populations for AGF under current and future scenarios.

4.2.6 Conservation prioritization and simulating assisted gene flow

We compared the metric DAP with an existing measure known as the Shapley Index (Haake et al. 2008). Shapley scores prioritize populations based on evolutionary isolation and contribution to overall diversity based on pairwise differentiation, and several similar metrics are widely used for conservation management (Collen et al. 2011; Jetz et al. 2014). Here, we used the method outlined in Volkmann et al. (2014), which maximizes within-species phylogenetic diversity using a network approach implemented in the software NeighborNet (Bryant & Moulton 2004; Huson & Bryant 2006). We used linear regression to test for a relationship between the log transformed Shapley index and DAP. Subsequently, we perform analyses of variance (ANOVA), to test for a relationship between population DAP values for each environmental variable and the response of square root transformed total catkin count. Finally we apply rank aggregation in the *RankAggreg* R package (Pihur et al. 2009) to achieve a consensus prioritization strategy that maximizes population Shapley values, whilst minimizing ADI.

Finally, we assessed for each environmental variable, at each sampling location, which populations would display reduced deviation from our model of optimum allele frequency. This population has potential to be used as a source of genetic material for assisted gene flow. Where several suitable populations were identified within the confidence interval of our regression, we selected the location geographically closest to the recipient population, as this could be considered the most conservative provenance sourcing strategy (Boshier et al. 2015). With this approach, we demonstrated how adaptive genetic diversity would be theoretically redistributed to minimize DAP and improve the adaptive potential of this species.

4.3 Results

After quality control, RAD sequencing produced a total of 173,460,998 reads, of which 79.1% aligned to the draft *B. nana* genome. Subsequently 73.2% of aligned reads mapped to a single unique position and were retained. Three samples were excluded due to low coverage. After filtering for SNPs that occurred in a defined portion of populations and individuals, this resulted in a *SNP_env* dataset of 14,889 SNPs over 8,727 loci. To address the question on introgression, mapping of the same set of *B. nana* reads and 867,994,425 *B. pubescens* reads to the custom 12 species reference resulted in 58.4% and 82.12% unique alignments respectively. Filtering of *B. nana* based on SNP_int dataset criteria resulted in 180,086 loci, whilst filtering of *B. pubescens* retained 124,701 loci. Comparison of the two datasets identified 4,987 reference mapped SNPs across 2,214 loci that were variable in both species.

4.3.1 Genome scans for local adaptation

Bayescan identified 382 putative outlier SNPs with a relaxed false discovery rate or 0.2. We tolerated an elevated rate of false discovery, to ensure that loci under selection were excluded for the computation of the neutral covariance matrix in environmental association analysis. Environmental association analysis in Bayenv2 subsequently detected 267 highly significant locus-environment associations, encompassing 303 SNPs (Table 4.2), comparatively, 23 loci were detected with small populations included. It is notable that only six loci were in common between Bayescan and Bayenv2 detection methods, and Bayescan candidate loci in did not report significantly higher Bayenv2 BF scores compared to the dataset as a whole. It should be noted that some SNPs report multiple significant relationships, in subsequent analysis we retained only the strongest relationship.

Table 4.2 Environmental variables and the number of associated loci.

Variable	Grouping	Description	Retained	EAA Loci
AMTemp	A	Annual Mean Temperature	X	17
MTColdQ	A	Mean Temperature of Coldest Quarter		24
MTColdM	A	Min Temperature of Coldest Month		23
MTWarmM	B	Max Temperature of Warmest Month	X	2
MTWarmQ	B	Mean Temperature of Warmest Quarter		4
MDR	C	Mean Diurnal Temperature Range	X	71
ISO	D	Isothermality	X	11
APrec	E	Annual Precipitation	X	2
PWetQ	E	Precipitation of Wettest Quarter		2
PDryQ	E	Precipitation of Driest Quarter		4
PWetM	E	Precipitation of Wettest Month		2
PDryM	E	Precipitation of Driest Month		3
Pseason	E	Precipitation Seasonality		1
PWarmQ	E	Precipitation of Warmest Quarter		4
PColdQ	E	Precipitation of Coldest Quarter		3
Slope	F	Slope (derived from elevation)	X	7
MTDryQ	G	Mean Temperature of Driest Quarter	X	7
Tseason	H	Temperature Seasonality	X	1
ATempR	H	Annual Temperature Range		2
MTWetQ	I	Mean Temperature of Wettest Quarter	X	7
Aspect	J	Aspect (derived from elevation)	X	4
Elev.	-	Elevation		12
Lat.	-	Latitude		6
Long.	-	Longitude		48

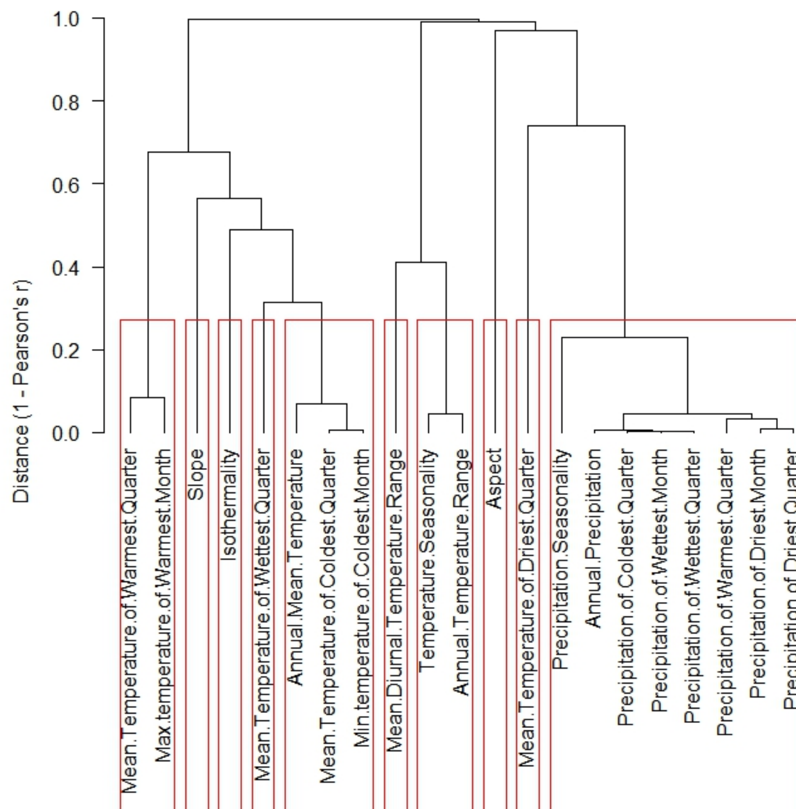


Figure 4.2. Topology of collinearity between environmental variables used in this study, at a threshold of 0.7. Slope and aspect were included and are not significantly correlated with any other variables. Table 4.2 details description of each variable name.

4.3.3 Introgression

Of 180,086 SNPs that passed filtering constraints, we found 4,987 that were shared between species and variable in both. STRUCTURE analysis revealed strong, but uneven evidence of bidirectional introgression between *B. pubescens* and *B. nana* (Figure 4.3). Individual estimates ranged from 0.003 - 0.081 proportion introgression into *B. pubescens*, and 0.002 – 0.044 into *B. nana* (Table S3). Overall, mean introgression from *B. pubescens* is estimated at 1.8%. Findings were highly concordant across independent runs, with a mean similarity score of 0.999. Of the 267 loci identified in the environmental association analysis, 49% were shared between species, compared to 36% in the dataset as a whole. Seven of these loci were also identified as having evidence of introgression. Spatial analysis of *B. nana* introgression found no significant association with latitude or altitude.

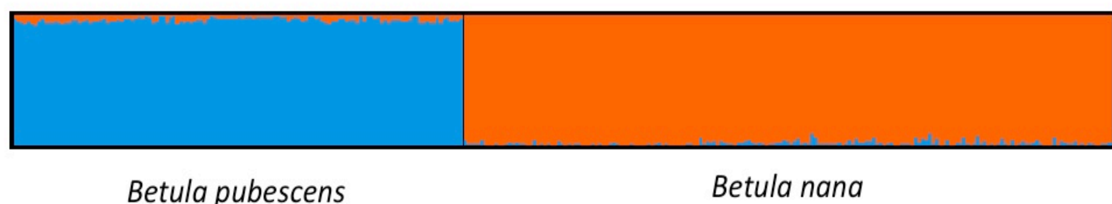


Figure 4.3 STRUCTURE plot identifying low levels of introgression between *B. pubescens* and *B. nana*.

4.3.2 Relationship between genomic and environmental data

As expected, geographic and ecological distance was highly correlated (r=0.35, p=0.002) throughout the study area. However all other Mantel and partial Mantel tests were found to be non-significant, with the exception of weak isolation by distance in putatively adaptive loci. Closer inspection revealed that this pattern is almost entirely attributed to two outlying populations at the southern extreme of the species range. These populations have previously been identified as potentially subject to severe genetic drift (Chapter 3), thus overall we conclude there is little evidence of isolation by distance or isolation by environment in these data. Comparing the number of locus-environment associations (Table 4.3) as a crude measure

of adaptive variable importance, we find a significant correlation with the percentage contribution of environmental variables defining species range in our MaxEnt model ($F_{1,8}$ = 5.55, p = 0.046), the notable outlier being maximum temperature of the warmest month (MTWarmM) for which far fewer associated loci were detected than might be expected.

Table 4.3 Contribution of retained environmental variables to the *Betula nana* species distribution model, and the number of environmentally associated loci.

Abbrev.	Environmental Variable	Correlated Variables	Percent contribution[1]	Jackknife importance	EAA Loci	EAA Loci (inc..cor)[2]
AMT	Annual Mean Temperature	MTColdQ, MTColdM	34.9	5.6	17	64
MTWarmM	Max Temperature of Warmest Month	MTWarmQ	22.1	32.2	2	6
MDR	Mean Diurnal Range	-	14.8	17.1	71	71
ISO	Isothermality	-	14.6	26.5	11	11
APrec	Annual Precipitation	PColdQ, PWetM, PSeason, PWetQ, PWarmQ, PDryM, PDryQ	7.3	8.0	2	21
Slope	Slope	-	2.8	3.4	7	7
MTDryQ	Mean Temperature of Driest Quarter	-	1.6	5.0	7	7
TS	Temperature Seasonality	ATempR	1.4	1.3	1	3
MTWetQ	Mean Temperature of Wettest Quarter	-	0.3	0.8	7	7
Aspect	Aspect	-	0.2	0.2	4	4

[1]Percentage contribution is calculated as the increase in regularized gain added to the contribution of the
[2]Ttotal number of SNPs associated with both the reported variable, as well as related highly correlated variables that were excluded from the MaxEnt model.

4.3.4 Gene ontology and RNAseq

The 267 loci were located on 185 unique scaffolds in the V4 dwarf birch genome. 177 retained flanking regions yielded successful blast hits with GO mapping results, and 166 yielded annotations. The most common species matches included *Vitis vinifera, Prunus* sp. and

Populous trichocarpa, of which the latter is closely related to *Betula* (Tuskan et al. 2006). Fishers exact test identified 24 enriched GO terms at a significance level of P<0.05 (Table 4.4).

RNAseq resulted in 17.4 million and 31.6 million reads for leaf and flower tissues respectively. The first 10bp of all reads were trimmed due to low quality. Overall 35 candidate regions showed evidence of gene expression in flower tissue (19%), 15 showed gene expression in leaf tissue (8%) and 13 showed gene expression in both (7%). We compared these to the proportion of loci in the overall dataset (8,727 loci located on 4,712 unique scaffols) with RNA matches and found that both flower (X^2 = 23.14, p<0.001) and leaf (X^2 = 8.59, p=0.003) expressed sequences are significantly over-represented in our putatively adaptive loci.

Table 4.4 Significantly overrepresented gene ontology terms identified in *B. nana* individuals.

GO-ID	Term	Category	P-Value	#Test	#Ref	Level
GO:0032550	purine ribonucleoside binding	F	6.76E-03	33	14	OVER
GO:0001883	purine nucleoside binding	F	6.76E-03	33	14	OVER
GO:0032555	purine ribonucleotide binding	F	6.76E-03	33	14	OVER
GO:0032553	ribonucleotide binding	F	6.76E-03	33	14	OVER
GO:0017076	purine nucleotide binding	F	6.76E-03	33	14	OVER
GO:0035639	purine ribonucleoside triphosphate binding	F	7.36E-03	30	12	OVER
GO:0097367	carbohydrate derivative binding	F	7.86E-03	34	15	OVER
GO:0032549	ribonucleoside binding	F	1.17E-02	33	15	OVER
GO:0001882	nucleoside binding	F	1.17E-02	33	15	OVER
GO:0006470	protein dephosphorylation	P	1.50E-02	7	0	OVER
GO:0004721	phosphoprotein phosphatase activity	F	1.50E-02	7	0	OVER
GO:0043412	macromolecule modification	P	1.93E-02	33	16	OVER
GO:0030554	adenyl nucleotide binding	F	2.20E-02	30	14	OVER
GO:0032559	adenyl ribonucleotide binding	F	2.20E-02	30	14	OVER
GO:0016788	hydrolase activity, acting on ester bonds	F	2.35E-02	16	5	OVER
GO:0005524	ATP binding	F	2.49E-02	27	12	OVER
GO:0044699	single-organism process	P	3.02E-02	66	43	OVER
GO:0006091	generation of precursor metabolites and energy	P	3.03E-02	6	0	OVER
GO:0032440	2-alkenal reductase [NAD(P)] activity	F	3.03E-02	6	0	OVER
GO:0006464	cellular protein modification process	P	3.55E-02	30	15	OVER
GO:0036211	protein modification process	P	3.55E-02	30	15	OVER
GO:0042578	phosphoric ester hydrolase activity	F	3.74E-02	8	1	OVER
GO:1901265	nucleoside phosphate binding	F	4.12E-02	38	21	OVER
GO:0000166	nucleotide binding	F	4.12E-02	38	21	OVER

4.3.5 Potential for adaptation and conservation prioritization

The deficit in adaptive potential (DAP) based on environmentally associated SNPs under present climate varied from 0.07 (SD±0.06) at Glen Cannich, to 0.39 (SD±0.24) at Beinn Enaiglair on the Western periphery of the species range. Mean population DAP was 0.19 (SD±0.08) (Table 4.5). BLUP estimates for MDR and MTColdM are presented in Figure 4.4 (and all estimates are given in Figure S8). Under future climate scenarios (which excluded slope and elevation analyses) mean DAP increased, from 0.22 (SD±0.10) to a maximum of 0.27 (SD±0.11) in scenario RCP8.5 (Table S4), but was not uniform across populations with many increasing (e.g. LX,PC) and others declining (LR).

Table 4.5 Deficit in adaptive potential across all BLUP analyses for present time environmental variables.

Pop	AMTemp	MDR	ISO	MTColdM	MTWetQ	MTDryQ	MTColdQ	Slope	Elev.	Combined	SD
BL	0.287	0.188	0.216	0.374	0.349	0.021	0.079	0.091	0.011	0.180	0.137
MO	0.056	0.232	0.237	0.117	0.167	0.251	0.156	0.043	0.018	0.142	0.089
BE	0.636	0.483	0.460	0.672	0.135	0.222	0.637	0.230	0.015	0.388	0.243
LH	0.059	0.069	0.480	0.307	0.009	0.390	0.057	0.131	0.006	0.168	0.178
BW	0.049	0.283	0.318	0.017	0.020	0.080	0.028	0.050	0.003	0.094	0.119
ME	0.126	0.135	0.081	0.078	0.013	0.383	0.220	0.015	0.018	0.119	0.120
GC	0.053	0.017	0.077	0.082	0.064	0.196	0.006	0.157	0.013	0.074	0.065
DE	0.136	0.159	0.353	0.048	0.114	0.373	0.323	0.144	0.012	0.185	0.133
AS	0.392	0.085	0.144	0.347	0.020	0.559	0.267	0.843	0.032	0.299	0.273
BB	0.363	0.469	0.300	0.337	0.056	0.146	0.516	0.106	0.005	0.255	0.184
PC	0.092	0.058	0.071	0.166	0.044	0.297	0.036	0.115	0.001	0.098	0.089
AV	0.242	0.317	0.348	0.464	0.014	0.268	0.443	0.080	0.016	0.244	0.172
MD	0.329	0.125	0.078	0.377	0.061	0.090	0.554	0.162	0.006	0.198	0.182
SL	0.009	0.117	0.111	0.090	0.033	0.102	0.067	0.065	0.032	0.070	0.039
MU1	0.213	0.311	0.561	0.090	0.012	0.536	0.066	0.027	0.009	0.203	0.220
MU2	0.142	0.314	0.550	0.036	0.089	0.161	0.063	0.566	0.003	0.214	0.215
LG	0.031	0.050	0.164	0.031	0.076	0.292	0.027	0.226	0.001	0.100	0.103
LL	0.012	0.149	0.230	0.026	0.005	0.026	0.161	0.075	0.025	0.079	0.081
BG	0.389	0.204	0.050	0.041	0.063	0.769	0.230	0.039	0.026	0.201	0.246
LR	0.076	0.125	0.141	0.051	0.003	0.530	0.020	0.013	0.001	0.107	0.167
RW	0.254	0.258	0.109	0.309	0.200	0.509	0.099	0.072	0.002	0.201	0.153
RB	0.138	0.160	0.402	0.253	0.063	0.155	0.159	0.112	0.007	0.161	0.113
LX	0.234	0.170	0.561	0.220	0.325	0.455	0.410	0.046	0.001	0.269	0.186
EM	0.383	0.292	0.479	0.128	0.175	0.756	0.109	0.084	0.008	0.268	0.238
SA	0.552	0.317	0.540	0.541	0.023	0.464	0.109	0.107	0.011	0.296	0.235
TD	0.304	0.211	0.149	0.558	0.174	0.550	0.491	0.038	0.008	0.276	0.212

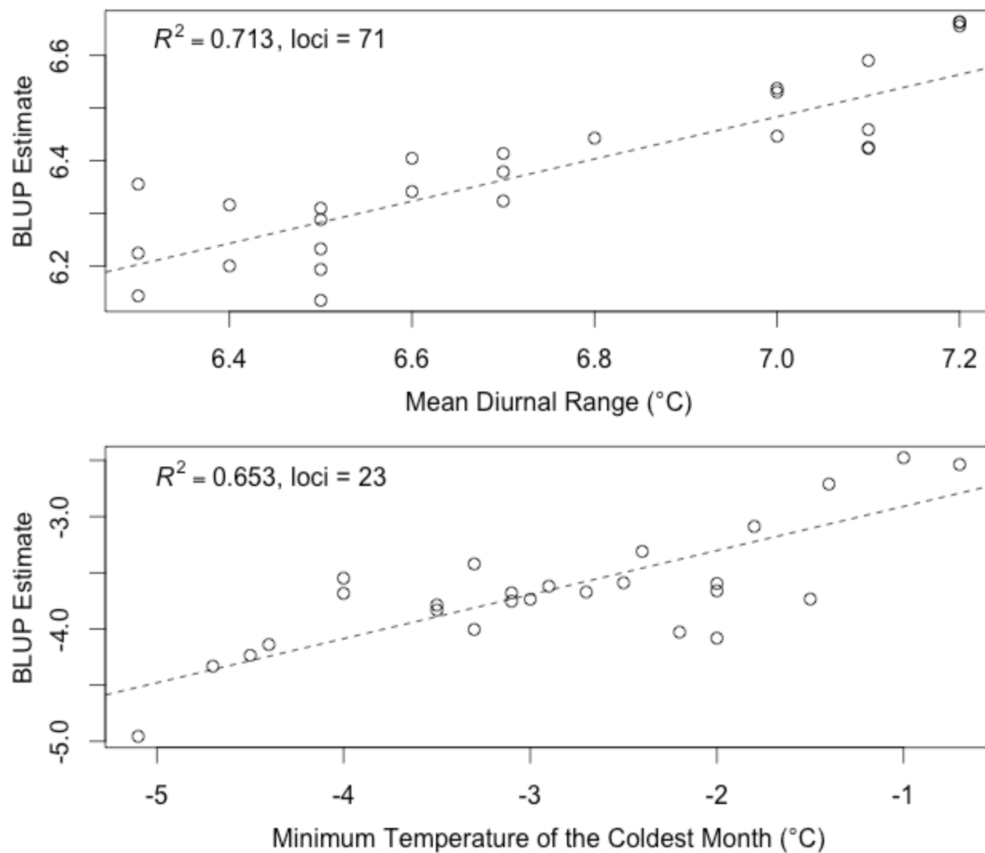


Figure 4.4 BLUP estimates for two environmental associations identified in the Bayenv2 analysis. Dotted line denotes proposed optimum genotype across the range of environmental values.

Population Shapley values are plotted in Figure 5A (raw plots are presented in Figure S7). Confirming our interpretation of DAP, we found that Shapley values were correlated with DAP for neutral loci ($F_{1,24}$=5.895, P=0.02) but not for putative adaptive loci (P= 0.93) (Figure 5B). In comparison with phenotypic data, we found that the DAP for annual mean temperature was significantly negatively correlated with mean population catkin counts ($F_{1,21}$=5.84, P=0.025) Figure 5C. Germination rate was similarly best explained by the interaction between Annual Mean Temperature and Mean Diurnal Range ($F_{1,11}$=7.13, P=0.022). This suggests that the lower the deficit in adaptive potential, the closer the population is to the optimal genotype, resulting in greater reproductive output. A consensus population ranking based on Shapley and DAP values is reported in Table S5.

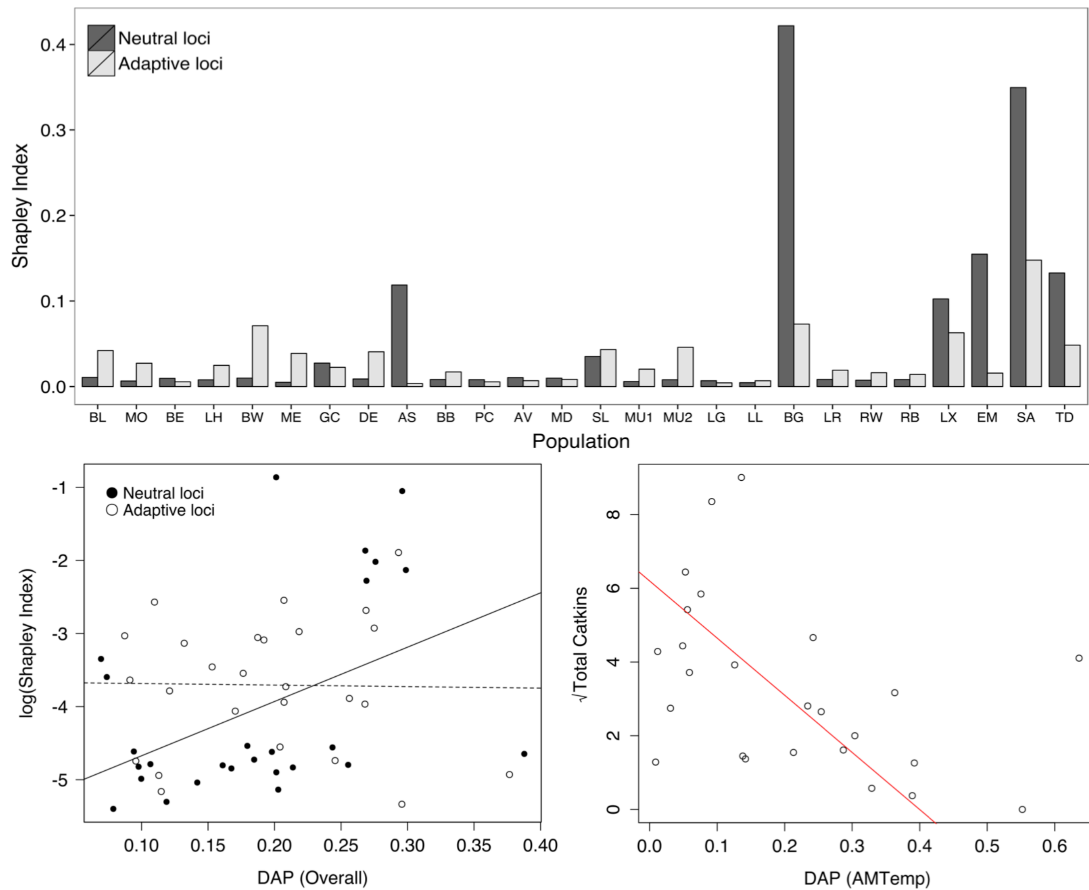


Figure 4.5 A) Barplot of Shapley index based on adaptive and neutral loci across UK *B. nana* populations ordered by latitude with northernmost populations to the left. B) Correlation of adaptive (solid line) and neutral (dotted line) with DAP. C) Regression of total catkin count against DAP for Annual Mean Temperature.

4.3.6 Simulating assisted gene flow

For each population across each environmental variable we identified the geographically closest 'donor' population with an allele frequency that would display reduced DAP (within confidence limits) at the 'recipient' site (plotted here for the environmental variables AMTemp and MDR, Figure 4.6, and for all variables, Figure S9). Overall, we most frequently find a pattern of dispersal from the center of the distribution towards the periphery, particularly at the Southern range edge, though there are exceptions such as transfer from the Northern to Southern range edge (e.g. MTColdQ, Figure S9). For all variables, we find some populations in which the genetic-environment relationship is within the confidence limits of our model, thus assisted gene flow from an alternative 'donor' population would not further reduce DAP. Thus

some surveyed populations, frequently in the center of the species distribution are successfully maintaining adaptive potential.

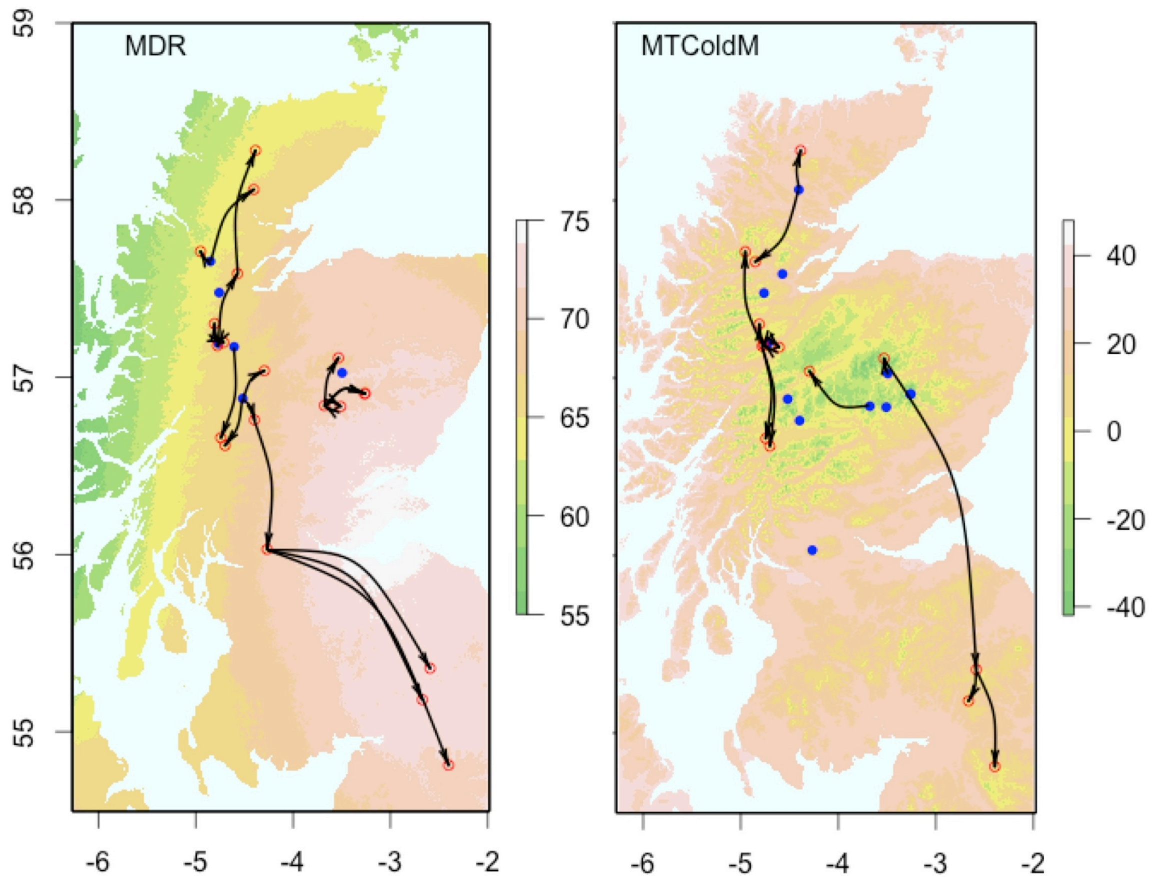


Figure 4.6 Hypothetical plots of assisted gene flow for dwarf birch in the UK. Arrows denote movement from donor to recipient populations (red circles). Blue populations report an allele frequency close to predicted optimums, thus do not require introduction of novel diversity. Base maps show MDR and MTColdM environmental values as a coloured raster grid.

4.4 Discussion

We present an approach to measure genotype-environment concordance in wild dwarf birch populations using thousands of genome wide SNPs, by estimating the adaptive allele frequency deviation from a theoretical genotype-environment optimum that varies across the species' range. We make the assumption that maximum population fitness is found when population allele frequency corresponds to the value predicted by a linear relationship of allele

frequency and environmental variable. Furthermore, we propose that populations exhibiting greater deviations across several genotype-environment associations are likely to be less well adapted relative to other populations. We provide initial evidence of this link by correlating our measure of the deficit in adaptive potential (DAP) with reproductive output, and by showing that candidate loci are significantly overexpressed compared to a random genome sample. If this approach is validated through robust reciprocal transplant experiments, we demonstrate a hypothetical assisted gene flow strategy for composite provenancing in order to maximize the adaptive potential of the various populations.

4.4.1 Identifying important environmental variables

Distribution modeling projects that the decline of dwarf birch across the UK is likely to continue and become increasingly severe, with almost total range loss possible by the end of the century under worst-case scenarios (Chapter 2). We overcame a classic caveat of SDMs linking models and fitness proxies by demonstrating that catkin production and seed germination rates are significantly higher at locations identified by the SDM as more favorable. Our predictions are consistent with observations in Svalbard, whereby apparent deviation from thermal requirements reduced dwarf birch germination rates (Alsos et al., 2003).

In comparison of SDM and EAA analyses, we found that four environmental variables contribute substantially (>10%) to the dwarf birch SDM (Table 4.3) as well as accounting for the largest proportion of adaptive loci (Table 4.2, 4.3). As such, there was significant agreement between SDMs and EAA for identifying important explanatory variables, adding further evidence that SDMs detect biologically meaningful variables that drive or influence adaptation. A similar result was reported recently between ranked environmental factors for habitat characterization and EAA (using a different association method) in Rellstab et al. (2016) for *Quercus* in Switzerland, but the authors did not investigate further potential parallels with SDMs.

Despite substantial range overlap and relatively recent speciation (Wang et al. 2016) we find surprisingly low levels of introgression from *B. pubescens* into *B. nana*. Hybridisation has been the focus of considerable interest for its potential to mediate species range shifts through adaptive introgression (Suarez-Gonzalez et al. 2016; Hamilton & Miller 2015), however here we find little evidence that it is contributing to *B. nana* persistence at the range edge or lower altitudes where species interactions are most likely. We do however find that adaptive and expressed loci are more likely to be shared with the sister species *B. pubescens*, suggesting that the genomic architecture of environmental adaptation may be conserved.

4.4.2 Detecting environmental associations

Overall, we detected 267 locus-environment associations across 21 environmental variables. Such a number is comparable to that found in similar studies on the White-breasted Nuthatch and Barley (Manthey & Moyle 2015; Abebe et al. 2015). Environmental and geographic distance was correlated, but overall we found little evidence of neutral or adaptive isolation by environment or geographic distance in this study. This contrasts with many published examples of local adaptation (Orsini et al. 2013; Sexton et al. 2014; Manthey & Moyle 2015). However other examples did not necessarily investigate species that have experienced recent population declines potentially affecting patterns of local adaptation. This is particularly demonstrated by the exclusion of a seven individuals from populations subject to severe genetic drift, which dramatically enhanced the number of associations detected. Overall, the occurrence of highly significant genotype-environment concordance, in the absence of significant isolation by distance, suggests a role for adaptation in structuring fragmented dwarf birch populations across the UK.

As expected, we find the most substantial deviations from the estimated adaptive optimum are in small populations at the range edge, where populations may not be able to keep pace with climate change. The current lack of concordance between genotypes and environmental variables is likely to increase under faster rates of projected environmental change. DAP for

Annual Mean Temperature significantly predicted declining catkin production and the interaction Annual Mean Temperature and Mean Diurnal Range further explained germination rate. Thus if environmental change outpaces the rate of adaptive evolution, then dwarf birch populations are likely to decline in reproductive output. Interestingly, the adaptive lag may not increase uniformly across the species range, as climate projections indicate highly heterogeneous combinations of conditions in the future. For some populations this results in a non-linear increase or subsequent decrease in DAP at future time points (Table S4).

We tested our estimate of adaptive potential against a commonly used metric, the Shapley index and found that both are correlated (Figure 4.5). This suggests that populations with the highest inferred conservation value are those with the greatest deviation from optimum allele frequencies. Thus wrongly exploited, this information means that conservation prioritization based on neutral diversity may inadvertently favour poorly adapted populations because they display a high degree of unique variation due to drift. We also note that the Shapley index, when based on adaptive loci does not correlate with DAP. Overall, we propose an approach where populations with a low DAP and high adaptive Shapley index are prioritized, as this would maximize current adaptation, as well as future adaptive potential (Table S5).

4.4.3 A hypothetical model of assisted gene flow

Estimating the current and future adaptive potential of species has important implications in guiding seed sourcing for population management and restoration (Salmela et al. 2010; Havens et al. 2015; Aitken & Bemmels 2016).. With the assumption that increased population fitness is achieved when deviation is minimized from a linear relationship of allele frequency and environmental variable, we propose an approach to identify putative donor populations that possess adaptive alleles in the frequencies that would reduce DAP in a recipient population. We showed that a hypothetical assisted gene flow strategy would require a substantial rearrangement of genotypes, particularly from the center of the range towards the periphery in dwarf birch. Following the evaluation of variation in DAP, it offers the possibility

to identify genetic material that will enhance the rate at which populations approach an optimum genotype for the corresponding environmental conditions, independent of the distance to the source population and the 'local is best' paradigm (Jones 2013; Boshier et al. 2015; Havens et al. 2015).

As a result, there is growing interest in conservation strategies incorporating climate-adjusted provenance of individuals for ecological restoration and forestry (Prober et al. 2015; Alfaro et al. 2014). In practice, implementation of assisted gene flow is likely to take the form of composite provenancing, whereby genetic material from a combination of source populations is used (Breed et al. 2013; Hodgins & Moore 2016). AGF is advantageous, because it can introduce or increase the frequency of preadapted alleles that may allow more rapid adaptation to track changing climate, alleviate inbreeding depression, increase adaptive potential and in the process provide a demographic safeguard by augmenting population size (Hodgins & Moore 2016). We note that although here we propose assisted gene flow of genotypes suited to the prevailing environmental conditions in the recipient population, alternatively, based on predicted future climate (Table S5), future provenance matching could be performed. Although somewhat substantial interventions, we note that strategies to this effect are already being proposed (Potter & Hargrove 2012; Aitken & Bemmels 2016), though limitations have been observed, for example in adaptation to cold and pest tolerance (Williams & Dumroese 2013).

4.4.4 Limitations of using correlation methods to measure adaptive potential

In the analysis presented here we have formulated a hypothetical model to describe the relationship between the population allele frequency at an adaptive locus and the value of an environmental variable at that location. We note a number of important and potentially limiting assumptions made in the development of this model that must be considered and validated before there is any degree of justification for use as a tool for species management.

Fundamentally, we highlight that correlation (environmental association analysis) based studies can identify correlation between environment and genotype, but cannot be used to prove causation. In this case, the model assumes that maximum fitness is found when allele frequency corresponds to an optimum predicted by a linear relationship between environmental variables and population allele frequency. However we note that we have not defined or provided evidence for the mechanism by which this relationship could arise, and further note that we have not performed a comparison with the wide range of other plausible models or alternative environmental association approaches (Rellstab et al. 2015). Important non-linear associations may also remain undetected, though statistical treatments for these are likely to be incorporated in EAA in the future (Fitzpatrick & Keller 2015). Similarly, populations have not been measured under common environmental conditions or habitat suitability values. We propose a study design to test this below. We also highlight that this approach cannot account directly for phenotypic plasticity (although this may have a genetic basis (Anderson et al. 2011)), or the evolution of novel mutations. However in small populations this evolutionary potential may already be reduced (Chevin & Lande 2011).

There are several further limitations arising from the use of genome wide RADseq derived SNPs and the sampling design. Firstly, RAD sequencing identifies only a variable subset of the genome (Lowry et al. 2016) possibly missing other important adaptive loci (Harrisson et al. 2014) of large effect. This concern may be address to some degree by the rapidly decreasing costs of whole genome sequencing, and larger numbers of genome-wide markers. Second, there is a large degree of error with which population allele frequencies have been estimated, as a result of small sample sizes. This could be addressed by increasing sample size in the estimation of population allele frequency. Third, a number of individuals have been selectively removed on the basis of severe genetic drift, which has the effect of increasing the fraction of correlated markers. The justification for removing these individuals arises through inference from analyses in Chapter 3, though no objective tests have been performed to validate this

process, and the threshold is arbitrary. Fourth, we highlight that the mean of residuals is closely related to the specified significance level for retained correlations in EAA. Thus if a higher threshold is set, overall adaptive deficit will decrease in the study, (though we suggest that comparisons across populations within species are less compromised by this issue).

Considering the significant ecological risks associated with strategies such as AGF, extensive model testing and validation of model predictions should be performed prior to considering management implications arising from model predictions presented here. We propose that this should take the form of reciprocal transplant experiments whereby a range of populations are grown at each of a number of sites with a range of environmental conditions. This would ensure that populations of differing provenance are exposed to the same set of environmental conditions, a confounding variable in the *in situ* analysis presented here. It would also provide an opportunity to increase the accuracy of measured environmental variables. Thus under the null hypothesis of an absence of genetic effects on fitness in environmental association analysis, measured fitness proxies (e.g. reproductive output) should be the same for each population at each location. Given this null hypothesis, it is then possible to test the hypothesis that individual fitness within a population is related to the deviation of population allele frequencies from the proposed theoretical optimum defined by our model.

If validated, we further caution against interpreting the bioclimatic measurements described in this model as specifically relating to the putative adaptive loci in dwarf birch. Rather, the causal environmental variable influencing fitness may be unmeasured, but closely correlated with a variable included in this study. In the future, increasing availability and resolution of remotely sensed data similarly has the potential to improve detection of genotype-environment associations (Myneni 2016; Deblauwe et al. 2016), particularly to edaphic factors (Aitken & Bemmels 2016). From a practical perspective, it is also important to note that if validated, adopting a set of increasingly complex seed supply requirements is likely to have practical and economic consequences for the industry (Box 1), which may make meeting

planting targets sustainably and appropriately increasingly challenging (Whittet, Cavers, et al. 2016; Whittet, Cottrell, et al. 2016).

4.4.5 Conclusions and hypothesis testing

Defining the adaptive potential of populations as a criterion to inform selection of plant material for genetic rescue, composite provenance, assisted gene flow or species reintroductions is currently the subject of intense debate and interest (Bingham & Ranker 2011; Prober et al. 2015; Gibson et al. 2016); this is likely to increase in the context of environmental change (Aitken & Bemmels 2016). Here we suggest a novel, but untested approach to permit rapid assessment of deficit in adaptive potential in threatened wild populations, which may, after validation and consideration of other factors such as effective population size and browsing pressure, have applications in the management and restoration of *Betula nana*. We suggest that comparison with reciprocal transplant experiments is appropriate next step in this study to provide essential validation of this method.

BOX 1: Implications for native plant management in the UK

Choosing suitably adapted seeds for management and planting schemes is an important consideration in forestry (FCS 2004; Forestry Commission 2011). It is now a requirement for grant schemes and is a component of UK sustainable forestry guidelines. Where possible, seed stock is required to be ecologically adapted to the site and prevailing conditions; maintain genetic adaptation and fitness of present populations (and adaptive capacity); sustain genetic variation; maintain and restore gene flow and conserve patterns of genetic structure that reflect evolutionary history (FCS 2004). There is not yet a consensus on whether future selection pressures from climate change may exceed the ability of native tree populations to respond (Williams & Dumroese 2013).

However, as demonstrated in the well studied and commercially important species *Pinus sylvestris*, enhanced understanding of the evolutionary ecology and adaptive variation (in addition to neutral variation) of native plants is likely to benefit management and conservation (Salmela et al. 2010). The current seed source zones are given for *B. pendula* in Figure 4.7. Based on the genetic differentiation (Chapter 3) and environmental site conditions reported here, comparative seed sourcing guidelines with validation through reciprocal transplant and germination experiments could be developed for dwarf birch. In the future, guidelines such as seed sourcing zones, may need to be updated to reflect adaptive genetic diversity, particularly in species where environmental conditions vary substantially across the range.

BOX 1: Cont.

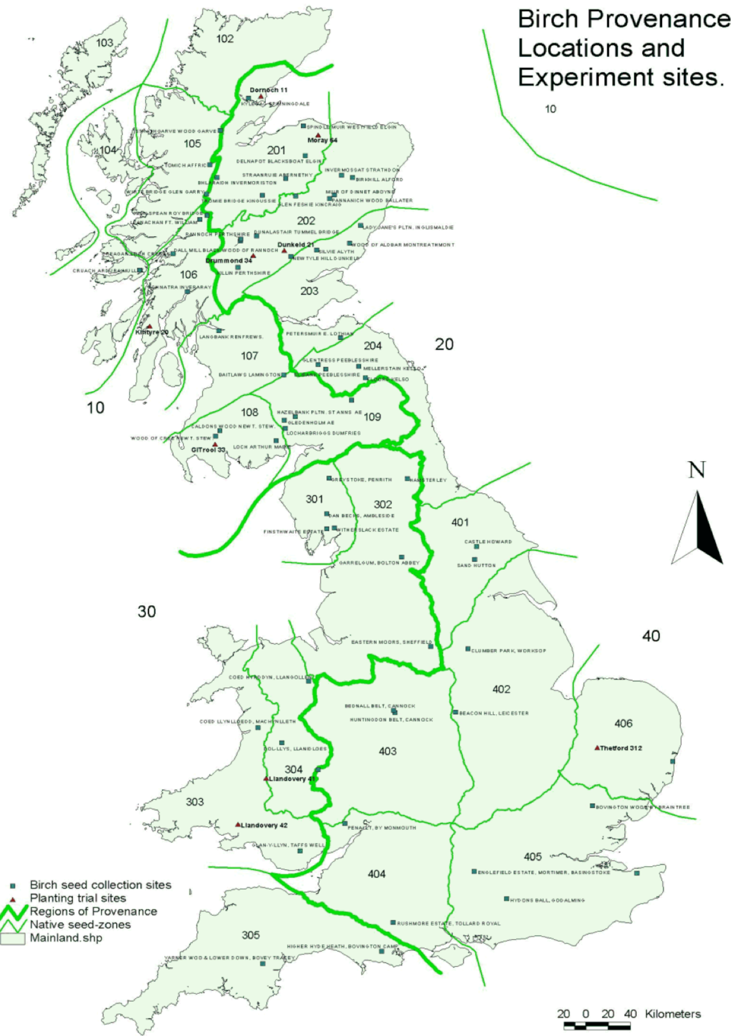


Figure 4.7. Birch provenance zones for *Betula pendula*. Adapted from Birch Group of the British and Irish Hardwoods Improvement Programme (2003).

Chapter 5: A review of the ecological genomics of *Betula*, and other montane/arctic species in Europe

Summary

The genus *Betula* is emerging as an informative system in which to study the post-glacial recolonisation of Northern Europe. Here we first review the observed patterns of population structure and genetic diversity in montane and arctic species with a range of life history traits including species with characteristics comparable to *Betula nana*. Second, we provide an overview of recent research elucidating patterns of structure and diversity of in genus *Betula*, with a particular focus on *B. nana* to provide context for the conclusions drawn from this thesis. Finally, we summarize the current and future activity for the conservation and restoration of *B. nana* populations in the UK, including surveys, propagation of seedlings and herbivore exclosures. We conclude with a summary of the key findings of this thesis that may be relevant to CASE partner organisations.

5.1 Comparative population structure and genetic diversity of montane trees, dwarf shrubs and perennial plants in Europe

The structure and diversity of populations is the result of recent and past demographic and population genetic processes including mutations, recombination, drift, selection and migration (Hewitt, 1999). In Europe, ice ages, climate refugia and subsequent post-glacial expansion have played a substantial role in shaping presently observed patterns (Huntley & Birks 1983). The different evolutionary history, phenotypic and life history characteristics of plants are likely to mean these patterns differ moderately or substantially among species (Lascoux et al. 2004), particularly due to persistence in putative 'northern refugia' (Schönswetter et al. 2005; Bhagwat & Willis 2008), dispersal ability (Feurdean et al. 2013) or the extent of gene flow among refugial populations (Lascoux et al. 2004; Cheddadi et al. 2006).

Improvements in high throughput genotyping has permitted an increasing number of larger scale reviews and analyses across single and multiple species (Lascoux et al. 2004; Meirmans et al. 2011; Eidesen et al. 2015). These show the generality of the pattern of 'Northen purity, Southern richness' first proposed by (Hewitt 1999), with the addition of frequently observed higher diversity due to northern admixture zones (Petit et al. 2003; Lascoux et al. 2004) (See Introduction). Overall, a comparison of studies is likely to improve understanding of how genetic structure and diversity is expected to vary across taxa with different life history traits. This may support identification of past refugia (Newton et al. 1999), highlight species with reduced diversity, or those likely to be susceptible to loss of genetic diversity as a result of isolation or fragmentation (Provan et al. 2008; Leonardi et al. 2012) and support conservation of plant genetic resources. Indeed analysis of the evolution and distribution of genetic diversity in several tree species has supported management and conservation both in the UK and Europe, for example *Pinus sylvestris* (Salmela et al. 2010), *Malus sylvestris* (Cornille et al. 2013) and *Fagus sylvatia* (Piotti et al. 2012).

Genetic studies in *Fraxinus* and *Pinus* found both tree species are characterized by high genetic diversity, but widespread haplotypic uniformity due to extensive gene flow and admixture (Heuertz et al. 2004; Robledo-Arnuncio et al. 2005; Cheddadi et al. 2006). Whereas the shrubs *Frangula alnus* and *Corylus avellana* display lower levels of genetic diversity and higher genetic differentiation in recolonized areas (but as expected, not in refugia) (Hampe et al. 2003, Persson et al. 2004). In this thesis, we focus on *Betula nana*, a species that may be considered approximately phenotypically intermediate between woody montane trees and shrubs, with wind dispersed pollen, relatively short-lived seeds, low growing stature, a large root to aerial shoot ratio and a degree of clonal reproduction. Together with species of *Salix* and *Juniperus*, in the UK and Scandinavia it forms montane scrub habitat at and above the tree line (Gilbert & Di Cosmo 2009), which itself exhibits conditions intermediate between the forest zone and the true arctic or alpine zones (Gilbert 2010).

The postglacial pattern of genetic diversity and population structure of *B. nana* is also apparently intermediate between montane trees and shrubs. *B. nana*, has characteristics of both, with relatively high genetic diversity as well as substantial genetic differentiation in the UK. In Scandinavia, the high level of genetic differentiation is not evident, suggesting that this pattern in *B. nana*, as well as in other shrubs may be due to anthropogenic and landscape factors. The near-total range contraction of *B. nana* in central Europe means it is not possible to directly ascertain refugial population structure and diversity, though inference from small population samples suggest genetic diversity is high (Dąbrowska et al. 2006), but not greater than in other areas of the range. Based on patterns observed in other montane species, it is plausible that several larger and more Northern refugia were present for *B. nana* with a high level of genetic diversity maintained through subsequent population isolation, with low levels of gene flow. In contrast, for less cold-tolerant species with refugia further to the South, colonization is frequently characterized by single or few haplotypes due to founder effects,

with other haplotypes often highly localized in the Southern portion of the distribution (Palme et al. 2004a).

Assessment of other small montane plants (Table 5.1) shows that many display lower diversity and higher differentiation, including *Primula, Rannunculus, Draba*, and *Calluna*, indicating that small plant size and limited dispersal potential may contribute to the higher degree of population structure found in these species (Glover & Abbott 1995; Rendell & Ennos 2002; Schönswetter et al. 2003; Skrede et al. 2009). Though broad overarching conclusions are difficult to draw, as there are often exceptions. For example the lichenized ascomycetes *Flavocetraria cucullata* and *F. nivalis* both display high diversity and low differentiation suggesting that in a pattern comparable with trees, despite their size, they achieve long distance gene flow.

Further comparison of *B. nana* population structure (Chapter 3) to other montane trees in Europe (Table 5.1) finds that anthropogenic disturbance over recent centuries may have had a similar impact on post glacial population structure. Research by Bacles et al. (2004) found an almost identical pattern to *B. nana* in *Sorbus aucuparia*, with remnant fragmented populations still supporting high levels of genetic diversity, that were only marginally lower than those in unfragmented European populations. Although they authors found no evidence of a bottleneck due to fragmentation, they suggest this may be due to it being too recent to detect. Here, in Chapter 3, we do detect a bottleneck associated with fragmentation in *B. nana*; our ability to detect this is possibly attributable to greater resolution (higher density genetic markers), and inference that the bottleneck may be older and perhaps more severe (see section 3.4). In another study analogous to work presented in Chapter 3, Persson et al. (2004) identified lower levels of within-population diversity at the range edge of *Corylus avellana*, potentially attributed to genetic drift in small isolated populations. Interestingly, although not explicitly tested, some of the small populations reported here are of a similar size to small populations in which severe genetic drift was detected in *B. nana*.

Overall, in contrast to the general pattern of 'Northen purity, Southern richness', *Betula*, *Calluna* and *Salix* species display patterns that differ somewhat, which Lascoux and authors (2004) suggest could be explained by the existence of more scattered northern refugia than for other species. It is surprising that the southern-most lowland *Betula* species used in this analysis (*B. pendula*), displayed this pattern, and through comparison to the literature we suggest this pattern would be more pronounced with the inclusion of *B. pubescens* or *B. nana* (Maliouchenko et al. 2007; Th.Thórsson et al. 2010). Persistence at higher latitudes likely means that more genotypes contributed to genetic diversity of the presently observed northern populations, with less severe bottlenecks and founder effects potentially also playing a role (Naydenov et al. 2007; Tzedakis et al. 2013).

An increasingly important research theme in the context of plant genetic diversity and population structure is the assessment of the prevalence of local adaptation. For temperate plants in Northern Europe, the majority of research has focused on forest trees (Savolainen et al. 2007; Salmela et al. 2010; Savolainen et al. 2013), with the degree of local adaptation determined by the balance of gene flow and the strength of selection. In a meta-analysis encompassing trees, shrubs, perennial/annual, clonal/non-clonal and self-compatible/incompatible organisms, Leimu & Fischer (2008) concluded that local adaptation appears to be independent of plant life-history traits, the degree of spatial and temporal habitat heterogeneity and geographic scale of the study. Rather population size had the largest and clearest effect, raising doubts for the ability for small plant populations to respond to climate change.

In one of the largest studies assessing neutral diversity in 27 high-alpine species, Meirmans et al. (2011) concluded that the most important determinant of a species' genetic structure is dispersal strategy. They observe that strong population structure only occurs when both seed and pollen dispersal are restricted. This is consistent with earlier work in *Fraxinus excelsior,* which found that pollen and particularly seed dispersal can play an effective role –potentially

even an enhanced role - in maintaining gene flow in a chronically fragmented landscape (Bacles et al. 2006). Nevertheless on the largest scales, major geographical features were consistent obstacles to gene flow for a range of species groups, despite high dispersal potential in some arctic plants. From a conservation perspective, higher levels of population differentiation appears to become more common in smaller, more poorly dispersing plants. Where effective population sizes remain sufficiently large (Chapter 3), population structure has the potential to maintain partitioned genetic diversity, though it is vulnerable to demographic decline. Thus for montane shrubs and perennials, more attention may need to be given to conserving a representative sample of populations, rather than in temperate trees where little diversity is partitioned between populations and overall census size is more important.

Table 5.1 Comparison of population structure and genetic diversity among montane trees, dwarf shrubs and perennials.

Species	Region	Pop	Dispersal	Marker	Findings	Author
Montane trees						
Sorbus aucuparia	Scotland	8	Insect-pollinated, bird-dispersed	Isozyme and cpDNA	Authors found remnant populations maintained high levels of diversity, that were only marginally lower than those for non-fragmented European populations. They found no evidence of a bottleneck accompanying fragmentation, but highlight that it may not recent enough to be detected.	Bacles et al. (2004)
Fraxinus excelsior	Europe	36	Wind-pollinated, wind-dispersed	SSRs	Population structure was characterised by low levels of inbreeding and differentiation, however the authors observed greater structure in proposed refugia and uniformity with higher diversity in northern more recently colonised area including the UK.	Heuertz et al. (2004)
Pinus sylvestris	Iberian Penins.	13	Wind-pollinated, wind-dispersed	cpSSRs	The authors found high within-population and low among-population genetic diversity, with overall haplotypic diversity very high. The research across an altitudinal gradient suggests effective vertical migration in response to changes in climate.	RobledoArnuncio et al. (2005)
Pinus sylvestris	Europe	106	Wind-pollinated, wind-dispersed	Mitochondrial DNA and cpSSRs	Genetic survey identified high diversity and low levels of population differentiation based on chloroplast markers. Mitochondrial diversity was lower with higher levels of differentiation. Overall population structure identified differentiation with Iberian populations, and relict populations in the Alps. Overall the authors conclude that the data support expansion in Northern Europe from a single glacial refugium, or gene flow between several refugium, with little subsequent differentiation as a result of recent fragmentation.	Cheddadi et al. (2006)
Pinus sylvestris	Scotland and Europe	21 and 8	Wind-pollinated, wind-dispersed	16 nuclear sequence loci	Relative to mainland Europe, Scottish pines did not show the expected reduction in nucleotide diversity due to colonization, though there was a reduction in rare nucleotide variants and substantial differentiation between regions. Within Scotland, Western populations showed slightly reduced nucleotide diversity compared to Southern and Eastern populations. The authors conclude that there is evidence for a relatively recent bottleneck and admixture of populations.	Wachowiak et al. (2011)
Betula sp.					See Table 5.1	
Dwarf shrubs						
Juniperus communis	Ireland	19	Wind-pollinated, animal-dispersed	Chloroplast AS-PCR and nuclear SSRs	The authors find substantial levels of differentiation based on chloroplast and nuclear markers, despite the expectation of high levels of gene flow in this wind dispersed species. They conclude that gene flow is restricted particularly over large geographic scales.	Provan & Bennett (2008)

Corylus avellana	Europe	40	Insect-pollinated, animal-dispersed	14 Allozymes	The authors identified lower levels of within-population diversity at marginal range edge populations. They also find evidence of a postglacial expansion bottleneck, and genetic drift in small isolated populations at the range margin, based on high levels of among-population genetic diversity. Finally, the authors report 996 unique multilocus genotypes among the 1055 stems comprising the European material, thus 59 samples (5.5%) were clones, suggesting an important role for vegetative reproduction within populations.	Persson et al. (2004)
Frangula alnus	Europe	78	Insect-pollinated, bird-dispersed	18 cpDNA, RFLP	The authors found a high degree of population differentiation and phylogegraphic structure. There was a strong contrast between the haplotype-rich glacial refugia and an almost completely uniform area of postglacial colonization. The temperate distribution consists of one very widespread and six highly localised haplotype. The authors suggest that Bird-mediated seed dispersal has permitted a rapid postglacial expansion of *F. alnus* as well as subsequent regular seed exchanges between populations across the large temperate portion of the range.Conversley, populations in the south of the refugia, are highly differentiation with unique genetic diversity, and appear to have contributed relativly little to the postglacial recolonization of temperate regions.	Hampe et al. (2003)
Perennual Herbs						
Primula scotia	Scotland	14	Insect-pollinated, wind-dispersed	Isozyme and RAPD	The authors identify an almost total lack of variation in this endemic insect pollinated plant, even in larger populations.	Glover & Abbott (1995)
Ranunculus glacialis	Europe	84	Insect-pollinated, wind-dispersed	AFLP and cpDNA	The authors identified substantial variation in population genetic diversity, low diversity in Scandinavian populations and the highest values in the Alps. They attribute this pattern in the North to bottlenecks during the pleistocene.	Schönswetter et al. (2003)
Campanula thyrsoides	Switzerland	32	Insect-pollinated, wind-dispersed	5 SSRs	No indication of genetic erosion or recent bottlenecks were found despite the species rareness and isolation of some of its populations. Population size did not affect molecular diversity in *C. thyrsoides* despite the theoretical prediction that small populations might lose genetic variation due to genetic drift, founder effects and population bottleneck.	Aægisdóttir et al. (2009)
Orthilia secunda, Monotropa hypopitys	Europe	35 and 19	Insect-pollinated, wind-dispersed	cpDNA nuclear ITS and SSR	Comparative phylogeography of two related plant species with overlapping ranges found that boreal *O. secunda*, has retained greater genetic diversity potentially as a result of persistence of more northerly populations outside of refugial areas at the LGM. With future northward range shifts projected, extinction of declining rear edge populations is likely to have a greater impact on *M. hypopitys*.	Beatty & Provan (2011)

Draba. subcapitata, high-arctic; *D. nivalis*, arctic to arctic–alpine; *D. fladnizensis,* arctic–alpine	Europe	83	Mainly self-pollinated, wind-dispersed	AFLPs and SSRs	The authors observed low within-population diversity and large between-population differentiation in all three species. In an analysis of three circumarctic species, fewer private alleles and less geographic structure was found in the northern-most species suggesting effective long distance dispersal, and a recent bottleneck. Two, less hardy, and more southerly distributed species had greater structure, most likely characterised by less effective dispersal, as well as a less severe postglacial bottleneck.	Skrede et al. (2009)
Meum athamanticum	Europe	23	Insect-pollinated, wind-dispersed	AFLPs	The authors unusually found high levels of genetic diversity both to the North and South of the Alps, the putative refugia. However to the North genetic differentiation is lower between populations, and there is no significant pattern of isolation by distance. By contrast to the South, populations are characterised by long-term isolation and little admixture of populations.	Huck et al. (2009)
Calluna vulgaris	Europe	22	Insect (and wind?) pollinated, wind-dispersed	RFLPs and SSRs	The authors found lower chloroplast diversity and differentiation in Northern Europe than in the South, however overall there was little evidence that haplotypes were geographically restricted. The authors also found that seed dispersal appeared comparatively more important than pollen dispersal for this species, based on G_{ST} values.	Rendell & Ennos (2002)
Other						
Flavocetraria cucullata and *F. nivalis*	Circumpolar	90	Wind dispersed (?)	Nuclear ITS	Authors find weak population structure and genetic differentiation in widespread arctic lichenized ascomycetes suggesting evidence of long distance gene flow. High levels of genetic diversity in the Arctic indicates long-term survival at high latitudes.	Geml et al. (2010)
Multispecies reviews						
27 high-alpine species	Europe	44-137	-	AFLPs	The authors focus on a set of 27 high-alpine species, and report strong population structure was found only when both seed and pollen dispersal was restricted. A second factor was soil preference, which determined distribution in refugia during the last glaciation.	Meirmans et al. (2011)
22 trees and shrubs	Europe	16-25	-	RFLPs	The authors find overall higher differentiation at southern latitudes with the highest diversity at admixture zones in intermediate latitudes. The authors also highlight that the observed pattern in *Betula* and *Salix* are not consistent with the other studied species, and could be explained by the existence of more scattered northern refugia than for other species. They draw comparisons to the observed populaton structure in *Calluna vulgaris*. Analysis also identified the same chloroplast haplotype for *B. pendula* present in Scotland and Scandinavia.	Palme et al. (2004a)

17 arctic species (seven herbs, seven dwarf-shrubs, two shrubs and one tree	Circumpolar	911	-	AFLPs	The authors predict that post-glacial admixture zones with higher genetic diversity should be most common in arctic species with particularly high dispersal capacity. With a focus on dwarf shrubs, the authors found similar geographical patterns in several species including *B. nana*, with five main groups across the circumpolar distribution, with the Ural mountains, Greenland and Arctic Ocean being the strongest barriers to gene flow.	Eidesen et al. (2015)

5.2 Comparative population structure and genetic diversity of genus *Betula* in Europe

The population structure of *Betula* species in Europe has been the subject of a considerable number of studies over the past two decades (Table 5.2), generally with the aim of addressing questions on colonization history as pioneer species after the last glacial maximum (Wang et al. 2014; Maliouchenko et al. 2007; Palmé et al. 2003), and investigating landscape patterns of introgression (Thórsson et al. 2001; Palme et al. 2004b). Other studies have focused on phylogenetic relationships within the genus (Li et al. 2007; Wang et al. 2016).

Population genetic research has largely focused on the tree birches *Betula pubescens* and *B. pendula*, which are the most abundant and widespread species. *Betula pendula* has considerable commercial value (Hynynen et al. 2010) and is subject to genome sequencing projects in both Finland and China (*unpublished*). The tree birches have also been studied in the context of plant-herbivore interactions (Olofsson et al. 2009; Hanhimäki et al. 1994; Dongen et al. 1994; Wesołowski & Rowiński 2006). Whilst the majority of early population genetic studies utilized a small numbers of chloroplast restriction fragment length polymorphisms (RFLPs) (Palmé et al. 2003; Palme et al. 2004a), with subsequent work employing chloroplast and then nuclear microsatellites (Wang et al. 2014; Tsuda et al. 2016a), the broad scale landscape genetic patterns have remained remarkably consistent.

As early as 2003 Palmé et al. (2003) reported extensive haplotype sharing across populations and low levels of population structure in *B. pendula* chloroplast DNA. Further work (Palme et al. 2004b) found that this was also the case in *B. pubescens*. In a small-scale comparison with the inclusion of *B. nana*, the authors reported that the majority of chloroplast haplotype variation in *B. pendula* and *B. pubescens* was within populations (F_{ST} = 0.26 & 0.32), however most *B. nana* variation was between populations (F_{ST} = 0.58) suggesting a higher degree of differentiation. Larger scale studies across Europe and Eurasia (Maliouchenko et al. 2007; Eidesen et al. 2015; Tsuda et al. 2016b) utilizing larger numbers of chloroplast and nuclear markers similarly found the lowest levels of population structure in the lowland *B. pendula*, with relatively low levels in *B. pubescens* and a higher degree of structure in *B. nana*. For example Maliouchenko et al (2007) reported moderately different genetic structure in the two tree birches, despite extensive haplotype sharing, with seven major groups in *B. pubsecens* and two in *B. pendula*. However in comparison to other broadleaved tree species, overall genetic structure was low, which they tentatively attribute to extensive dispersal of small birch seeds. In a recent study of the tree *Betula* species in the UK (Wang et al. 2014), for which a small number of samples contributed to this thesis, population structure was detected in sampled *B. nana* populations, but not in *B. pendula* or *B. pubescens*. Similarly, significant isolation by distance was detected in *B. nana* and *B. pubescens*, but not in *B. pendula*.

Overall, several authors have concluded that low levels of structure in *B. pendula* are consistent with postglacial expansion from lower latitudes and few, if any, refugial populations further north. A slightly higher degree of structure in *B. pubescens*, and the persistence of unique but geographically localised haplotypes was attributed by Maliouchenko et al. (2007) to the limited existence of northern microrefugia, though it is not the only explanation for such a pattern. Finally, high levels of population structure in *B. nana* suggest the persistence of a greater number of northern refugial populations, which have subsequently contributed to contemporary landscape genetic patterns. Furthermore, for *B. nana* population substructure

appears to have led to the maintenance of moderately higher genetic diversity (Thórsson et al. 2010; Dąbrowska et al. 2006). This is consistent with ecological knowledge of the species, where *B. pendula* is characterized as a widespread lowland species, *B. pubescens* marginally preferring upland areas and wetter more acidic soils (Atkinson 1992) and *B. nana* the northernmost montane species restricted above the tree line. Data in this thesis support this inference with substantial genetic differentiation of UK *B. nana* populations in the UK. Higher levels of genetic diversity and lower differentiation in populations surveyed in Scandinavia, could be attributed to admixture from a larger number of refugia and subsequent colonization, which would be geographically plausible (Ibrahim et al. 1996). Though the lower levels of genetic differentiation suggest that geography and terrain, which are less heterogeneous in Scandinavia, may play a role (de Bello et al. 2013).

Table 5.2 Summary of phylogeographic studies on *Betula* sp. in Europe.

Species	Region	Pops	Markers	Summary	Author
B. pendula	N. Europe	9	Allozymes	In a two species comparison, the authors report slightly higher levels of genetic diversity for *Betula pendula*, and lower differentiation (F_{ST} = 0.032), in comparison with *Acer platanoides* (F_{ST} = 0.099).	Rusanen et al. (2003)
B. pendula	Europe	47	RFLP	Two very common haplotypes contributed to 84% of the total sample, although a number of rare haplotypes were highly localised to specific areas. The distribution of haplotypes suggest persistence of *B. pubescens* further north that the glacial refugia for many species, and rapid colonization Northwards from intermediate latitudes. The authors conclude that *B. pendula* displays a moderate level of between-population diversity compared to other studied trees. The authors highlight that the fixation index for *B. pendula* is in a similar range to that of *Calluna vulgaris* (heather), a species for which intermediate latuitude refugia are also proposed.	Palmé et al. (2003)
B. pendula *B. pubescens* *B. nana*	Scandinavia and W. Russia	21	Chloroplast RFLPs	The majority of variation in *B. pendula* and *B. pubescens* is within populations (F_{ST} = 0.26 & 0.32), however most B. nana variation is between populations (F_{ST} = 0.58). Overall the authors did not find significant phylogeographic structure in any of the three species. Furthermore, they found extensive haplotype sharing suggesting hybridisation and introgression, particularly at higher latitudes - though they highlight that this study focuses on chloroplast DNA perhaps inferring chloroplast capture, and they caution against extapolating to the nuclear genome.	Palme et al. (2004b)
B. nana	Poland	1	Nuclear RAPD	The authors report a higher than expected level of retained genetic diversity in a small relict population of *B. nana*. They attribute this to an initially large and diverse gene pool and frequent sexual reproduction, with diversity perhaps also being maintained by the long lifespan of clones.	Dąbrowska et al. (2006)
B. nana, B. pubescens, B. nana (limited)	Eurasia	53	Chloroplast RFLPs and SSRs	Similar number of haplotypes and within-populaton diversity in *B. pubescens* and *B. pendula*. Common haplotypes were shared among species. They found different genetic structure in the two tree birches, despite extensive haplotype sharing, with seven major groups in B. pubsecens and two in *B. pendula*. However in comparison to other broadleaved tree species, overall genetic structure was low, perhaps due to effective and extensive dispersal of small birch seeds. The authors suggest that the additional structure in *B. pubescens* could be attributed to different postglacial dynamics, with *B. pubscens* populations persisting at higher latitudes. Furthermore, a higher incidence of hybridisation with B. nana, owing to a larger area of sympatry could also have contributed to this pattern.	Maliouchenko et al. (2007)
B. pendula, B. pubescens, B. nana	Iceland, Greenland, Scandinavia and Scotland	12	Chloroplast RFLPs	The authors find high levels of introgression in the chloroplast genome between the three species, and this is more extensive than similar reports at lower latitudes in Europe. Despite clear introgression, *B. nana* shows greater sub-division and differentiation, and higher genetic diversity than *B. pubescens*. Population substructure appears to have led to the maintenance of this higher diversity. The authors suggest colonisation of Iceland from Northern Scandinavia.	Th.Thórsson et al. (2010)

B. pendula, B. pubescens, B. nana	UK	78	Nuclear SSRs	Genetic structure was detected in sampled *B. nana* populations, but not in *B. pendula* or *B. pubescens*. Similarly, significant isolation by distance was detected in *B. nana* and *B. pubescens*, but not in *B. pendula*. Furthermore, greater introgression was detected between the diploid *B. nana* and tetraploid *B. pubescens* than with diploid *B. pendula*, perhaps due to more extensive range overlap.	Wang et al. (2014)
B. nana and *B. pubescens*	Europe	59	Nuclear AFLPs and plastid DNA	*B. nana* markers support expansion from two major refugia in Eurasia, one south of and one east of the North European ice sheets. By contrast, *B. pubescens* displays a lack of geographic structure. The authors also find evidence for increasing hybridisation with latitude suggesting hybridisation has been most abundant at the postglacial expansion front, with backcrossing mainly to *B. pubescens*. For *B. nana*, European populations form a single population structure cluster.	Eidesen et al. (2015)
Six species (including *B. pendula, B. pubescens* and *B. nana*)	Eurasia	129	Nuclear SSRs	All species reported low levels of admixture likely as a result of introgression. Consistent with other studies, the authors found *B.nana-B.pubescens* introgression to increase with latitude. *B. nana* reported lower genetic diversity in marginal populations such as Iceland. Based on 18 loci across four diploid *Betula* species, *B. nana* reported the highest genetic differentiation (F_{ST}=0.053), and intermediate values for allelic richness and expected heterozygosity. These values were however lower than those for *B. pendula*.	Tsuda et al. (2016b)

5.3 Region wide population structure and genetic diversity of *Betula nana*

Until relatively recently, phylogenetic studies that include dwarf birch were limited, and centered on questions such as the origin of *B. nana* in Iceland (Thórsson et al. 2010) and the diversity of remnant populations in Poland (Dąbrowska et al. 2006). The assessment of *B. nana* population structure and diversity reported in this thesis for the UK and Scandinavia is the first high-resolution assessment of dwarf birch populations and builds on other published data for this species. It is also a larger, but overlapping dataset, with that published in Wang et al. (2014). Interpreted together with a recent detailed European wide study by Eidesen et al. (2015) this provides a valuable opportunity to interpret both local and regional population structure of *B. nana*.

Based on AFLP markers, Eidesen et al. (2015) identify that *B. nana* likely recolonized from two or three refugial areas; to the south of the main European ice sheet, to the East and possibly in Beringia. By contrast, *B. pubescens* shows evidence of just a single large refugia. Both species' population structure from Eidesen et al. is reproduced in Figure 5.1, showing that the samples presented in this thesis likely all originate from one phylogeographic group at the broadest scale. However it is notable that our higher resolution analysis identifies very clear differentiation between Scandinavia and the UK (Figure 3.3). The higher degree of diversity we observed in Scandinavia is also consistent with a possible contribution and admixture of genetic material from a second eastern refugia, whereas the UK may have been colonized from just the Southern refugia. With *B. nana* now extinct from the proposed refugial areas, it may be difficult to ascertain whether this pattern is correct with greater confidence.

The authors also note a weak correlation between latitude and introgression, and suggest hybridization between the two species has been most common at the advancing range edge. This is consistent with the pattern found in (Wang et al. 2014). Though there was evidence for introgression, we did not find a significant geographical pattern in this study (Chapter 4).

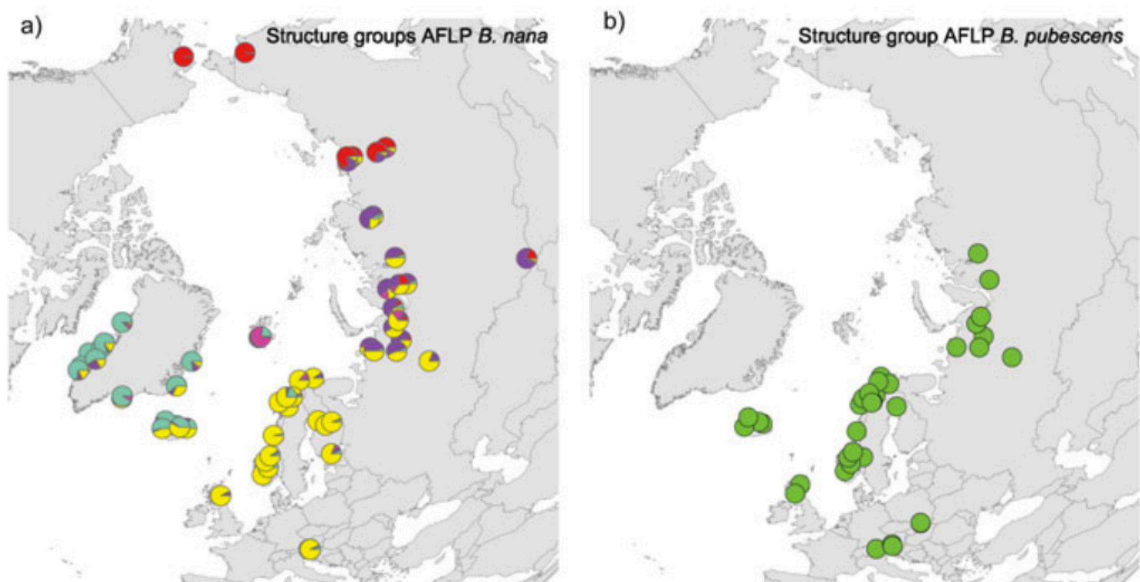


Figure 5.1. Geographical distribution of AFLP diversity for (a) *B. nana* and (b) *B. pubescens* inferred from STRUCTURE analysis in Eidesen et al. (2015).

5.4 Current and future work by CASE partners for the conservation of *B. nana* in the UK

Historical records collated by the National Biodiversity Network and several partner organisations (Chapter 2) report 2,245 observations of dwarf birch in the UK. However aside from the research presented in this thesis, only two detailed surveys and one research project have been undertaken:

The UK B. nana survey

The most recent was a limited survey in 2012 of the three surviving dwarf birch populations in Teesdale, Spadeadam and Emblehope areas of the UK, totaling four plants. Surveys of the surrounding area found one further population in 2015, consisting of ~3 individuals which was not included in this thesis. Genetic analysis by the author confirmed the number of individuals present at these sites, despite two locations (SA and EM), being characterized by single extremely large individuals that appear to have spread clonally. The conclusion of this assessment by Natural England and the Forestry Commission was that managers should first attempt propagation of the very small number of harvested seeds, with a long term strategy considering introduction of seedlings originating from Southern Scottish populations (pers comm. Tom Dearnley & Simon Webb).

The 2008 Dundreggan Estate survey

Prior to this, in 2008 CASE partners Trees for Life undertook a survey across a 21km area centered on the Dundreggan, Corrimony and Glenmoriston Estates. *Betula nana* was found in 31% of 491 plots, with large continuous areas above 500m a.s.l., particularly on Northern slopes. Between 450-500m, *B. nana* occurred in small scattered patches, and was absent below this altitude and on South facing slopes. Preference for Northerly aspects was similarly found across a larger study in this thesis (Chapter 2), however region wide assessment of altitudinal prevalence is difficult, because the abiotic conditions defining the tree line vary

across Scotland. Thus dwarf birch can occur as low as 150m a.s.l. in the West (e.g. population Ben Eniglair), and up to 800m in the Cairngorms (Korner 1998; Gilbert & Di Cosmo 2009). Based on the presence/absence on plots across the study area, the density of *B. nana* is mapped in Figure 5.2.

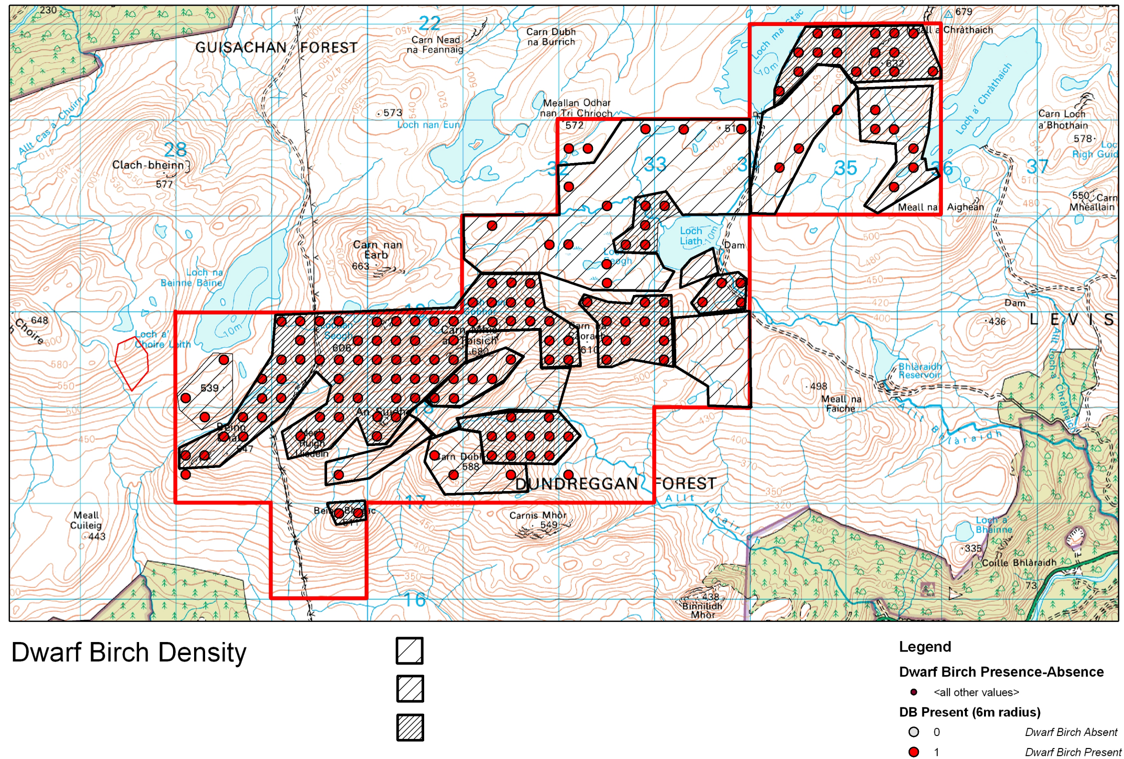


Figure 5.2. Distribution of *B. nana* across the Trees for Life study area. Red dots denote presence within the study plot of 6m radius. This survey was conducted in the summer of 2008 by Mark Richards published as a report to *Trees for Life*.

Similarly, analysis of plant size and catkin production across the survey identified a number of trends. Mean height varied from 6.6 - 33.4cm which is broadly intermediate in values observed in this thesis. Catkin production was found to be higher in taller plants that exceeded the height of surrounding vegetation. Similarly, catkin production was found to be reduced by even moderate levels of browsing pressure, and no catkins were found on plants with >80% of shoots browsed. Both findings are largely consisted with the geographically wider analysis presented in this thesis. The mean number of catkins produced per plant was 1.95 in the survey across Dundreggan estates (though we note some small differences in methodology),

which is below the average from populations reported in this thesis despite comparatively low grazing pressure, high genetic diversity and very high HSI identified in our analyses. This disparity may be attributable to annual variation in catkin production regionally or nationally due to climate and weather conditions, though this would require additional research to verify. Overall the data suggest that Dundreggan is a high quality habitat for *B. nana,* and future climate projections suggest it is likely to remain moderately favorable over the remainder of the century.

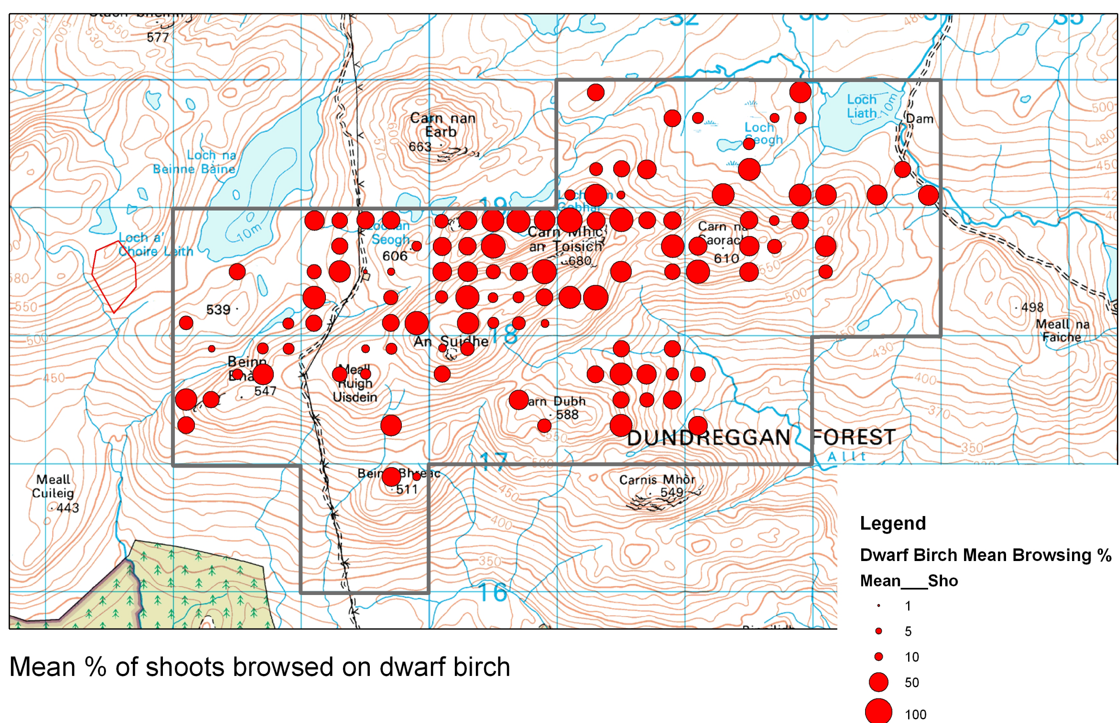


Figure 5.3. Mean *B. nana* browsing percentage across the Dundreggan study area. This survey was conducted in the summer of 2008 by Mark Richards published as a report to *Trees for Life*.

Regional assessment of montane scrub habitat

One further study focused on the interactions between climate and land use in driving the distribution of tree line scrub (montane scrub; including *B. nana, Salix myrsinites* and *Juniperus communis*) in Scotand (Gilbert 2010), for which the locations of some populations of *B. nana* were contributed to this thesis. This research made a number of findings, which are supported in further detail in this thesis. Firstly, the response to browsing was found to be complex and

difficult to characterize, but overall heavy browsing was likely to limit scrub expansion, consistent with our findings here.

Secondly, the author concludes that based on the spectrum of habitat conditions on which montane scrub species occur, there is potential for expansion of populations and species ranges – though this may be currently limited by land use, and in the future climate change. Interestingly, the author also reports that the Loch Muick population grows on a site that differs substantially in soil, aspect and slope properties to most other populations in Scotland and raises the possibility that it is a different strain. Here we have assessed the genetic diversity and differentiation of UK populations, and do indeed find that Loch Muick is differentiated, but not substantially more so than other populations. This supports the authors original conjecture that *B. nana* can tolerate a broader range of habitats than those it currently occupies. Taken together the work of Gilbert (2010) and the thesis presented here provide important insights to manage the conservation of *B. nana* and montane scrub habitat in the UK.

Propagation of *B. nana* seedlings

Current activity for the conservation of dwarf birch has two themes. The first is propogation of seedlings at the *Trees for Life* Dundreggan tree nursery. In addition to ongoing projects with *Betula pendula, B. pubescens, Juniperus communis* and *Pinus Sylvestris,* conservation managers have also begun large scale propogation and planting of *B. nana.* The need for this intervention was initially highlighted by Gilbert (2010) who found little evidence of natural recruitment in a subset of surveyed *B. nana* populations. All seeds are currently being sourced from the 2002 Dundreggan Exclosure (see population DE within this thesis). To date, >5000 propagated individuals have been introduced to four sites, including in Glen Cannich (original population analysed in this thesis GC), Glen Feshie (GF) and Glen Affric (not surveyed), as well as in the central and northern part of the Dundreggan estate. Although a limited amount of seedling

mortality has been observed (as expected), it is currently too early to assess the success of this planting strategy.

Based on the results of this thesis, it is interesting to note that population DE displays amongst the highest seed production of any measured population in the UK. Similarly, these plants also have a very high germination rate. Surprisingly, population DE was selected to provide source material out of convenience due to its close proximity to a 4x4 track, but this has inadvertently proved very advantageous and highly effective. We note that there is considerable genetic differentiation among UK populations (Chapter 3). We also propose that a portion of this may be attributable to local adaptation (but see caveats, in Chapter 4). If this is the case, it may be desirable to source seed from alternative locations for restoration across different parts of the UK range. For example, by matching elevation or climate attributes, rather than minimizing geographic distance (Salmela et al. 2010). Based on the preliminary germination experiments reported here conservation managers may find a substantially lower germination and survivability rate in these instances.

The importance of discouraging indiscriminate use of seed resources of geographically distant provenance, whilst also not necessarily restricting choice to the most local provenance was highlighted for Scotland nearly two decades ago (Ennos et al. 1998). The effective and appropriate selection of plant genetic material for reforestation and restoration continues to be discussed in the literature (Jones 2013; Boshier et al. 2015; Havens et al. 2015; Whittet, Cavers, et al. 2016), with further research necessary before local adaptation can be incorporated confidently into national management guidelines; such as Seed Sources for Planting Native Trees and shrubs in Scotland (FCS 2006).

Exclusion of herbivores

The second strategy is the erection of herbivore exclosures to encourage the regeneration of montane scrub species. These have generally been located at sites where several montane

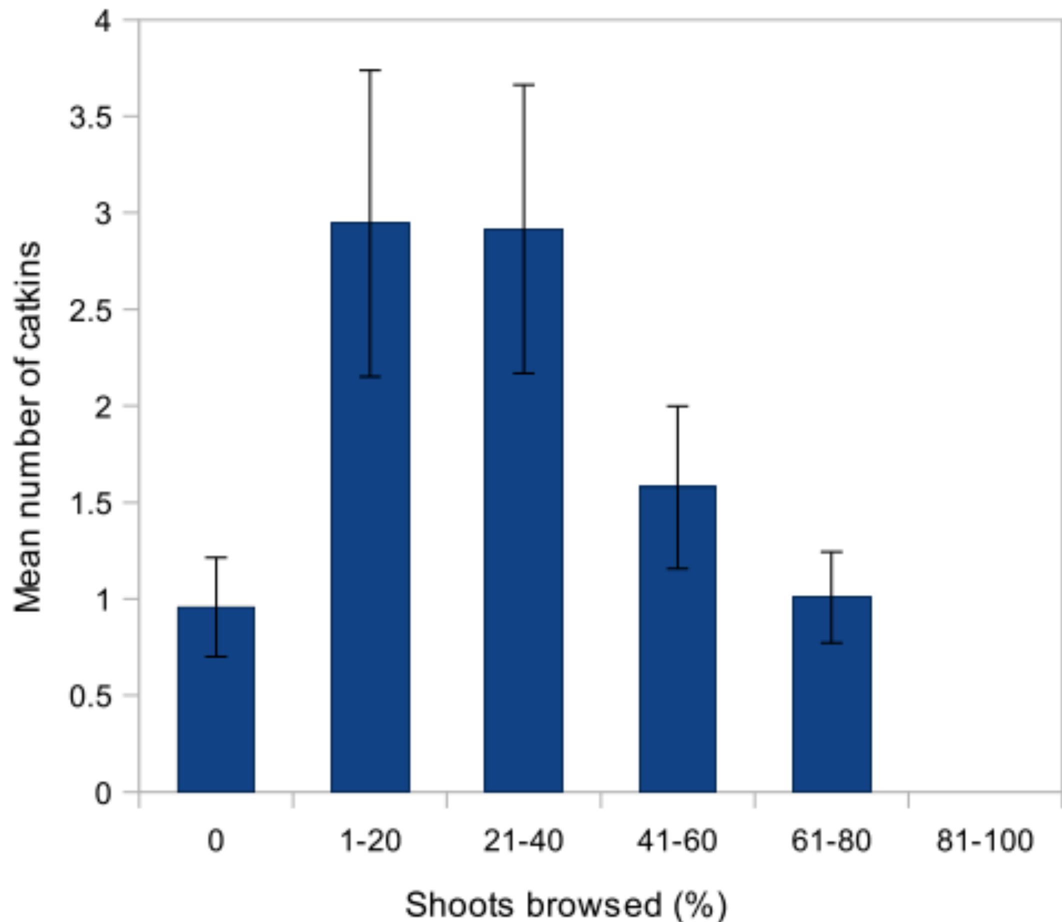


Figure 5.4. Mean number of catkins against percentage of shoots browsed. Figure reproduced from (Richards 2008).

scrub species are present, and there are sufficient plants to eventually form a dense scrub layer. One *Trees for Life* exclosure (Balnacarn) is sufficiently old to have had a repeat survey after 10 years. *B. nana* was found in 73 quadrats in 2008 compared to 42 in 1998. Though there may be some degree of sampling error due to changes in the methodology used, this suggests that *B. nana* may have become larger and/or spread to new quadrats in the past decade. Mean height also increased by ~10cm. In another exclosure, surveyed after 6 years, catkin production was measured and found to be a mean of 26.6 per individual inside the exclosure, compared to 1.95 outside. Together this data suggests that excluding herbivores may be the most effective short-term strategy to support the conservation of dwarf birch.

5.5 Conclusions and policy summary for CASE partners

1. Based on historical records, we confirm with more detail than previously known that dwarf birch has declined across the UK in recent decades. Analyses of historic introgression suggests that this is part of a broader long-term decline since the last glaciation.
2. Browsing appears closely linked to catkin production and population expansion as shown by this thesis and previous surveys on the *Trees for Life* Dundreggan Estate, thus reduction in browsing pressure may be the most effective strategy short-term strategy for the conservation of dwarf birch.
3. The habitat suitability of *B. nana* populations across the UK is likely to decline under future climate, however this decline is not uniform. Thus the identification of small regional refugia where the rate of change is more moderate for prioritization may be an effective medium-term conservation strategy. Chapter 2 provides an example of this approach based on populations included in this thesis, though it could be expanded to all records from the UK.
4. Genetic diversity in the UK is comparable to that in Scandinavia, which suggests that dwarf birch is resilient to some of the negative consequences of fragmentation and reductions in gene flow despite an order of magnitude decline in effective population size. We observe that severe genetic drift becomes apparent only at extremely small effective population sizes. Preventing decline of populations to single figures – where possible – is therefore of paramount importance. In populations with a census size below ten, genetic drift may substantially reduce the future conservation value of the germplasm at these sites. Unusual genetic variants at these sites may not be worthy of conservation even if they are unique, as they may be deleterious.
5. Here we have found a moderate to high degree of genetic differentiation across the UK, with some evidence of local adaptation. In the long term, for reintroduction and

restoration of *Betula nana*, we recommend diversification of seed source populations. Comparison of germination success and survivability from different stock would also further inform appropriate restoration strategies. Based on this, the pertinence of composite provenanceing and assisted gene flow could be evaluated in the future. Chapter 4 gives some hypotheses of advantageous assisted gene flow strategies that could be tested (though in addition to and not instead of other more conventional conservation measures).

6. Contrary to initial expectations, we do not find strong evidence for a deleterious impact due to introgression from other *Betula* species (similarly, we also do not find a clear advantage through adaptive introgression). Nevertheless, it may be prudent to avoid large-scale plantation *B. pendula* (and to a lesser extent, *B. pubescens*) near to sensitive *B. nana* populations. However we acknowledge with exclusion of browsers, a degree of regeneration of these tree species is likely to occur naturally. Specifically where sites are being managed for seed collection, it may be desirable to remove other Betula species, and/or avoid collection from putative.

Chapter 6: Conclusions

6.1 The contribution of this thesis

Here, I have presented an integrated assessment of the distribution and population genetic condition of dwarf birch in the UK. In collaboration, we have mapped and sampled all known dwarf populations in Britain, and populations over a similar area in Scandinavia. We have taken phenotypic, environmental and demographic measurements of each population. We have assessed genotypes using microsatellite and RAD-seq approaches. We have estimated past distributions from herbarium and other records. By analysing these data, we have: mapped past, present and predicted future distributions, elucidated demographic and colonization history since the LGM, assessed the impact of fragmentation, estimated levels of introgression and evaluated the degree of local adaptation to current and future climate. Together, this combination of approaches provides extensive information with which to inform future conservation management by CASE partner organisations. The ways in which we have combined and co-analysed our different datasets is innovative, and may provide wide applications for other species of conservation interest.

6.2 Novel insights for the conservation of dwarf birch

A number of important conclusions for dwarf birch have arisen from these analyses. We have shown unequivocally that dwarf birch has been declining in the UK over recent decades. In one of the first demonstrations of a correlation between niche model predictions and phenotypic fitness traits (but see Wittmann et al. (2016) where this has also been attempted), we have shown that MaxEnt models based on current climate are effective predictors of dwarf birch catkin production and germination success. If future climate models are accurate, then dwarf birch is likely to experience a further decline in habitat suitability in the UK, suggesting further reductions in fitness and consequential range contraction in the near future. However, the

declines in habitat suitability are not predicted to be uniform in severity across the species distribution, so there may be micro-refugia that are promising candidate sites for conservation. As dwarf birch is a key component of montane scrub habitat, it is possible that our predictions and conclusions for this species may apply to other less studied species that respond in a similar fashion or that overall community composition is likely to change.

We have established that the current landscape of reduced genetic diversity in the UK arose as a result of a recent population bottleneck coinciding with historic shifts in land use. Despite this, one of our most important insights is that dwarf birch populations show a surprising level of resilience to genetic drift until relatively small effective population sizes. Low levels of genetic drift have also been observed in other forest fragmentation studies, thus this may be an emerging characteristic of tree population genetics, montane species included (*sensu lato* Lowe et al. 2015).

We have also shown evidence of local adaptation in dwarf birch populations, with highly significant correlation of allele frequencies with environmental conditions. Using a novel analytical approach, we have estimated the degree to which each dwarf birch population in the UK deviates from optimum allele frequencies. These results can be used to identify the key environmental limiting factors and suggest sites and patterns of assisted gene flow that can remediate them. Reassuringly, we show that this novel method of assessing local adaptation (based on correlations of allele frequencies and environmental variables of occupied sites) identified similar environmental determinates of dwarf birch distributions as MaxEnt models (based on presence/absence species data correlated with environmental variables among all occupied and unoccupied sites). Consistency between these methods has not previously been reported in the literature.

We note that contrary to initial fears among CASE partners at the commencement of this study, introgression does not appear to be significant in the sample of populations analyzed. Nonetheless, it is possible that high levels of *B. pubescens* pollen could be limiting seed

production in some populations, and this was not explicitly tested. Overall, niche and range overlap is projected to increase, thus buffer zones and other strategies to reduce contact between species may be prudent.

In our studies, we found that small relictual populations of dwarf birch, mainly at the southern range edge, had signatures of a high level of genetic drift, and also had allele frequencies that did not appear well-adapted to their environments. However, under widely used neutral-marker based conservation prioritization approaches, these populations appear of high importance with a high level of evolutionary distinctiveness at neutral markers. We suggest that it may be prudent to include data from adaptive loci in conservation, and highlight that these gave different conclusions, at least in this study, to approaches based on neutral loci. We recommend that the distinctive southern-range edge UK dwarf birch populations are not good sources for assisted gene flow.

6.3 Methodological advances

This thesis has demonstrated a number of novel methodological advances: In chapter two I demonstrate the advantages of incorporating multiple independent markers sets with different mutation rates in population genetic analyses. Despite *in silico* methods for identifying and testing microsatellite markers, their development remains time consuming; thus here I have demonstrated a pipeline for extracting simple short microsatellite sequences from RADseq reads.

My results show that these perform almost as well as conventional microsatellites. It is anticipated that many researchers could extract these RAD-SSR sequences from existing datasets, thus providing additional genomic information for analysis. As the length of currently available RADseq reads already exceeds those used in this thesis, this method will provide increasing quantity and quality of data, as longer and more highly variable microsatellites with more complex repeat motifs can be extracted.

In chapter three I demonstrate a novel approach to assess the degree of putative local adaptation dwarf birch populations. Utilizing the development of high-density genome wide markers and environmental association analysis software, this method provides a population estimate of degree to which populations deviate from proposed optimum allele frequencies. We show that this can be applied for current environmental conditions, estimated under future climate and also as a method to inform assisted gene flow. Lowry et al (2016) highlight that current RAD marker densities are still insufficient to identify the vast majority of adaptive loci, but nevertheless we show that even with a limited number of loci the estimate of ADI does not vary substantially. Indeed as the number of reads generated from Illumina sequencing increases, it will be possible to analyze even higher density markers or resequenced genomes to further improve this estimate.

Scripts and instructions for implementing these approaches have been made available to the community via Github: https://github.com/JamesBorrell

6.4 New questions and future research

Characterization of adaptive loci and development as a model organism

Despite the availability of the *Betula nana* genome and a limited amount of expression data for two tissue types, it remains difficult to identify the functional sequence linked to the SNP loci with signatures of adaptation in our studies. This has been a challenge in many similar studies on local adaptation (Keller et al. 2012; Wilczek et al. 2014; Suarez-Gonzalez et al. 2016; Hecht et al. 2015). Where genes or expressed sequences can be identified, it is similarly challenging to then elucidate function and how precisely this relates to the correlated environmental variable. In the future *B. nana* may be highly suitable for further development as a model woody plant, on account of its small size, experimental tractability, short generation time and small genome. Others have already developed methods for genetic transformation of *Betula*

(Valjakka et al. 2000), and a high quality genome assembly is available (Wang et al. 2013). This would improve our ability to target and annotate loci linked to adaptation.

Empirical tests of composite provenancing in dwarf birch

The analysis presented in chapter four suggests that a population with allele frequencies close to the local environmental optimum should display higher fitness than those with a substantial deviation from the optimum allele frequency, however covariance with habitat suitability and population variability makes this challenging to empirically measure. Future research could apply assisted gene flow to assemble climate adapted and climate non-adapted populations at a series of trial locations. With this approach it would be possible to experimentally test the hypothesis: does the newly developed metric DAP predict reproductive output and other fitness related traits. Furthermore, this analysis could include a population of local provenance to compare the 'local is best' seed sourcing strategy that may indicate important additional factors such as epigenetics.

Identification of microrefugia

Based on MaxEnt models, a fruitful further research theme would be to identify locations where the magnitude of projected environmental change is minimized in the UK. Where dwarf birch is present, these may be important sites in which to prioritise conservation. Similarly, where dwarf birch is not present, but sites with high current and future habitat suitability are identified; these could be candidate microrefugia for reintroduction. Identification of the most important variables driving the distribution of dwarf birch (see species distribution modeling, chapter 2, and environmental association analysis, chapter 4), could facilitate this approach.

Supplementary Information

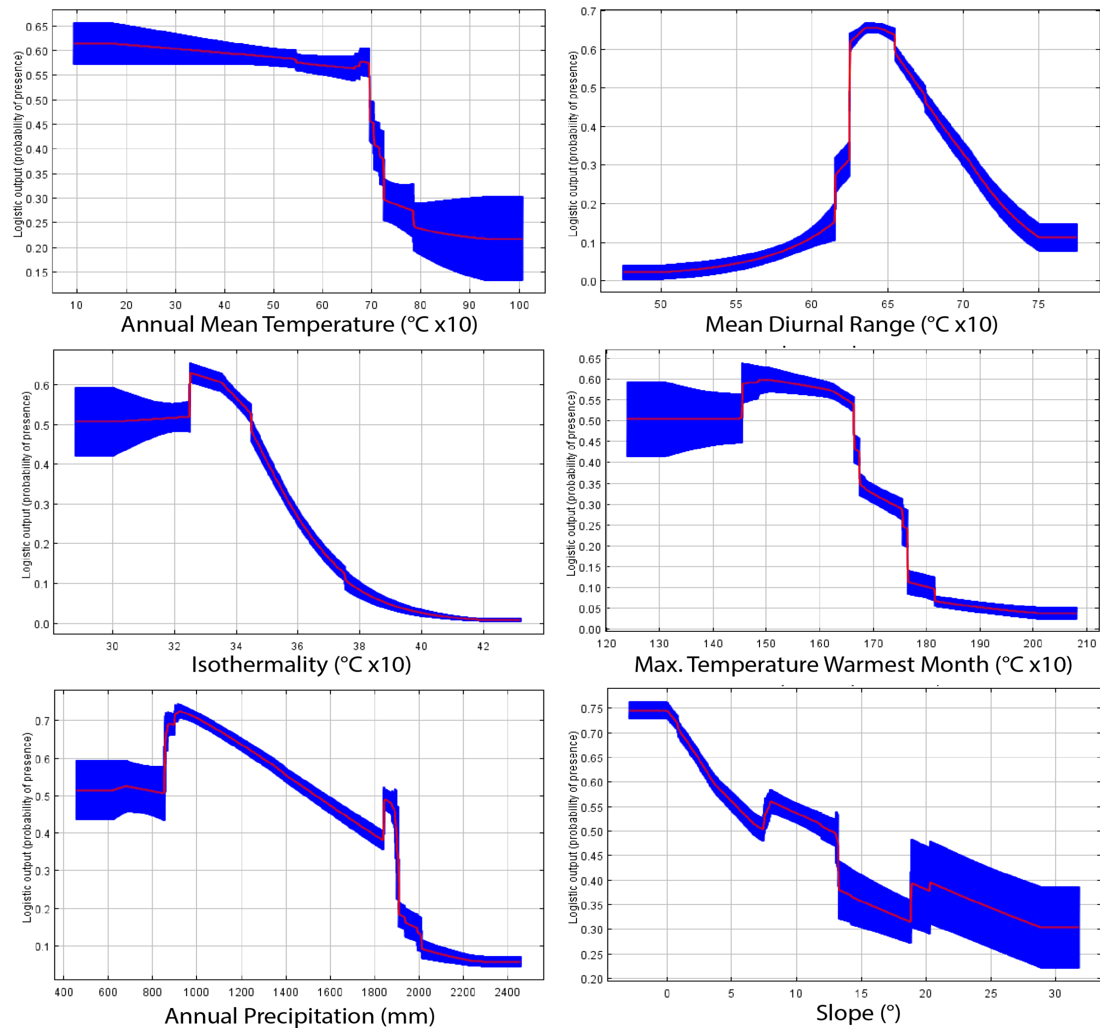


Figure S1. Species distribution model variable response curves for six environmental variables contributing substantially to the *B. nana* species distribution model. Note that temperate variables are multiplied by 10 in the bioclim database.

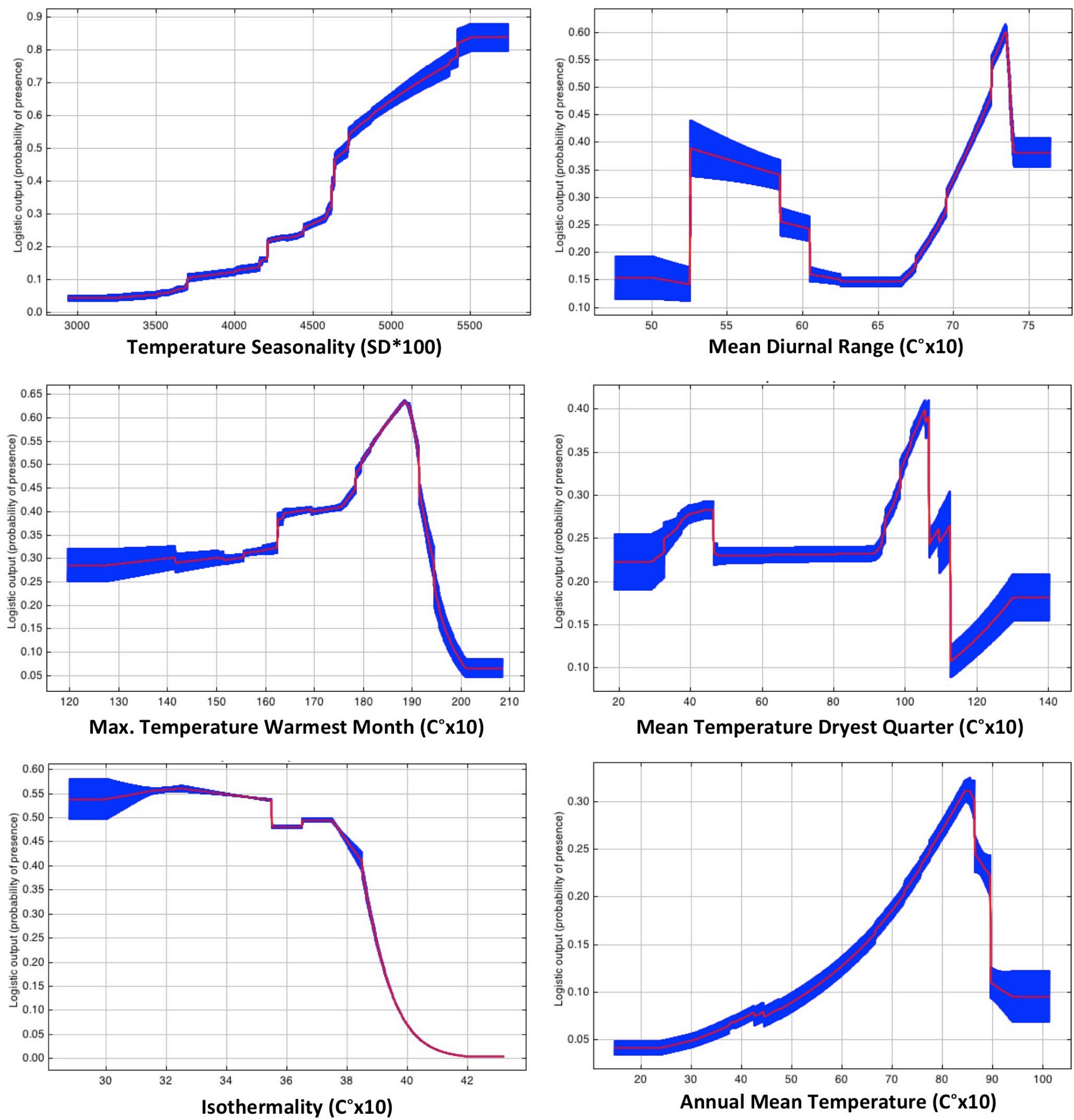


Figure S2. Species distribution model variable response curves for six environmental variables contributing substantially to the *B. pubescens* species distribution model. Note that temperate variables are multiplied by 10 in the bioclim database.

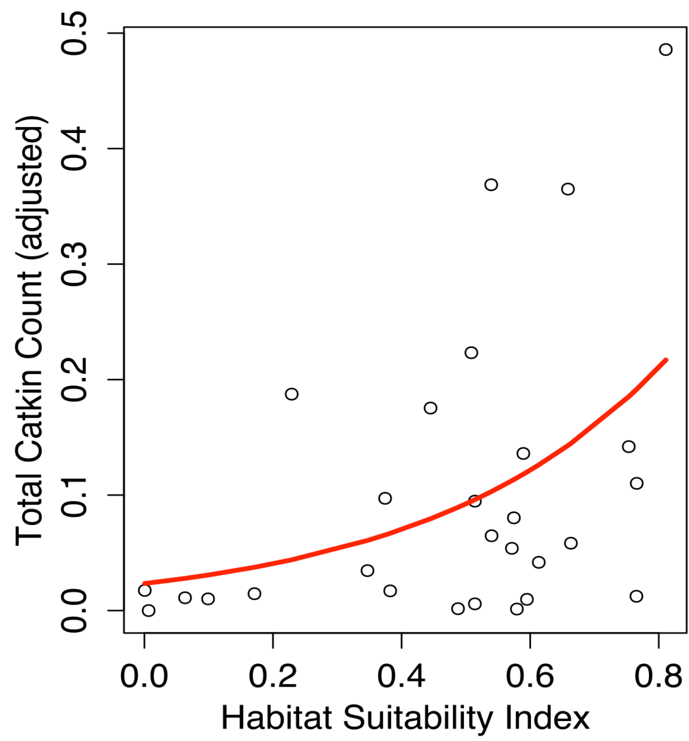


Figure S3. GLM of mean population total catkin count, corrected for plant area, against the MaxEnt derived habitat suitability index showing that more suitable sites are associated with a greater number of catkin and higher percentage germination.

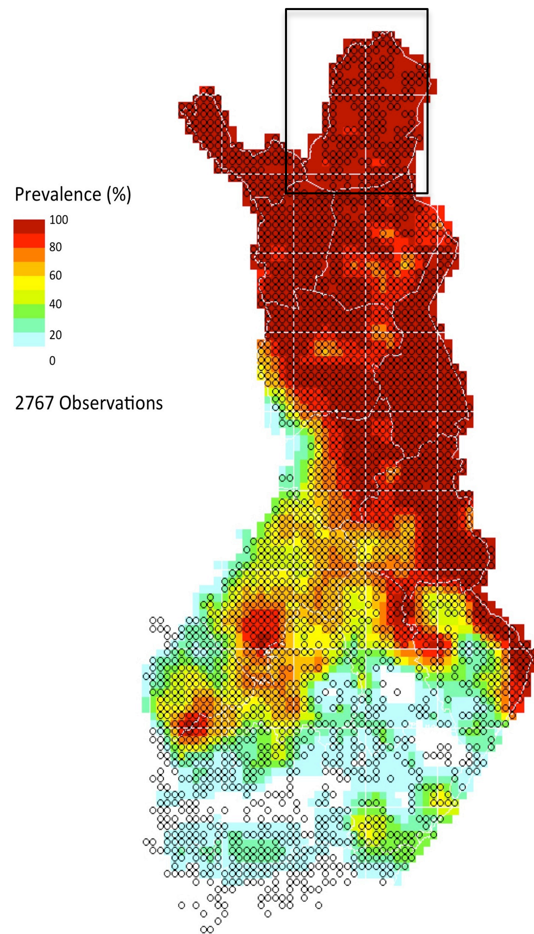


Figure S4. Prevalence of *B. nana* in Finland with rectangle denoting main study area. Figure is adapted fromLampinen, R. Lahti, T. 2016: Plant Atlas 2015 - University of Helsinki, Museum of Natural History, Helsinki. (www.luomus.fi/kasviatlas/)

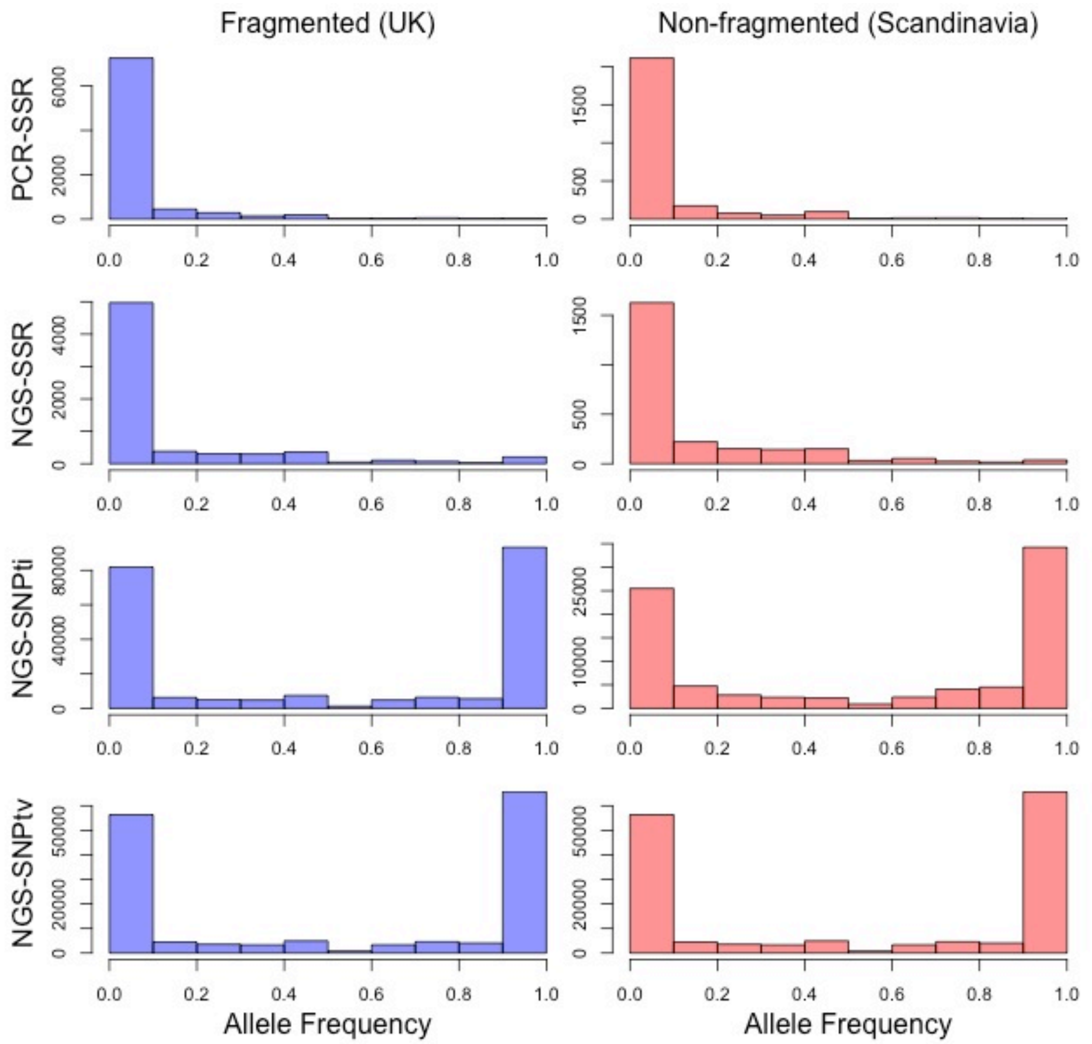


Figure S5. Allele frequency spectrum plots across all marker types and regions.

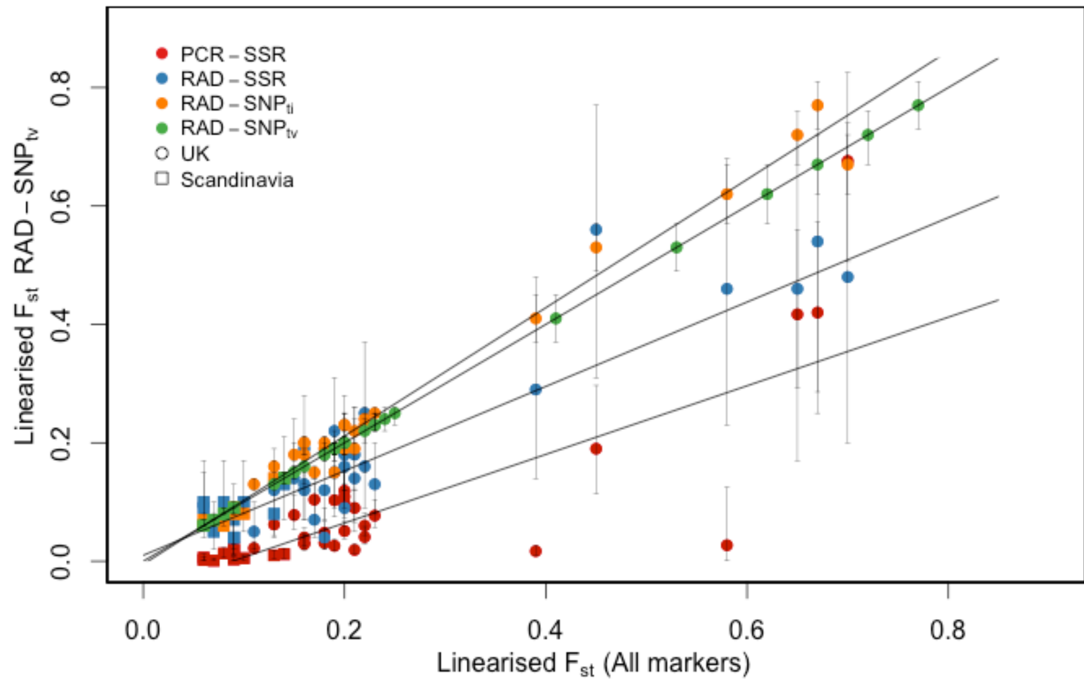
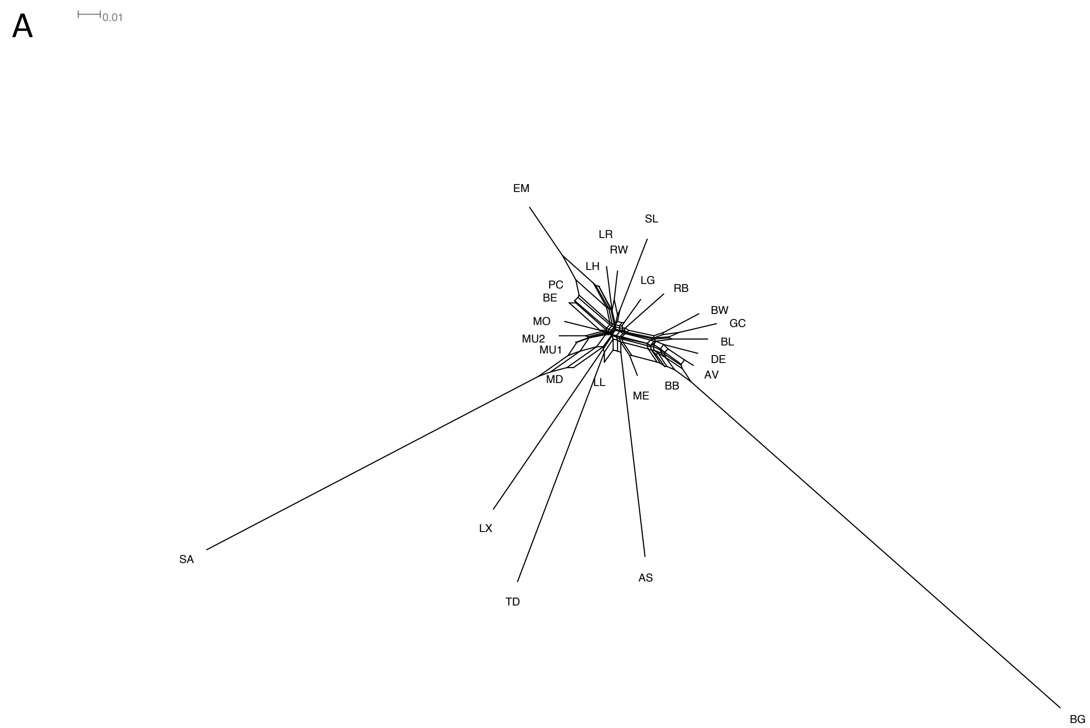


Figure S6. Comparison of Linearised F_{ST} values across marker types.



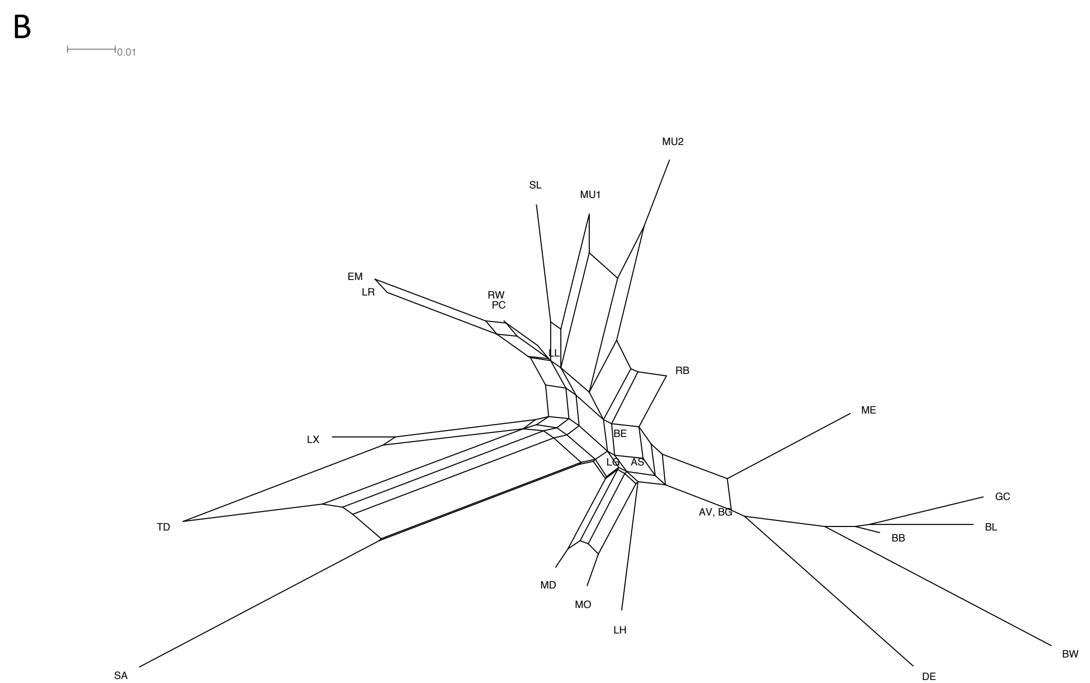


Figure S7. NeighbourNet networks of A) neutral F_{ST} and B) putative adaptive F_{ST} for 26 populations of dwarf birch in the UK.

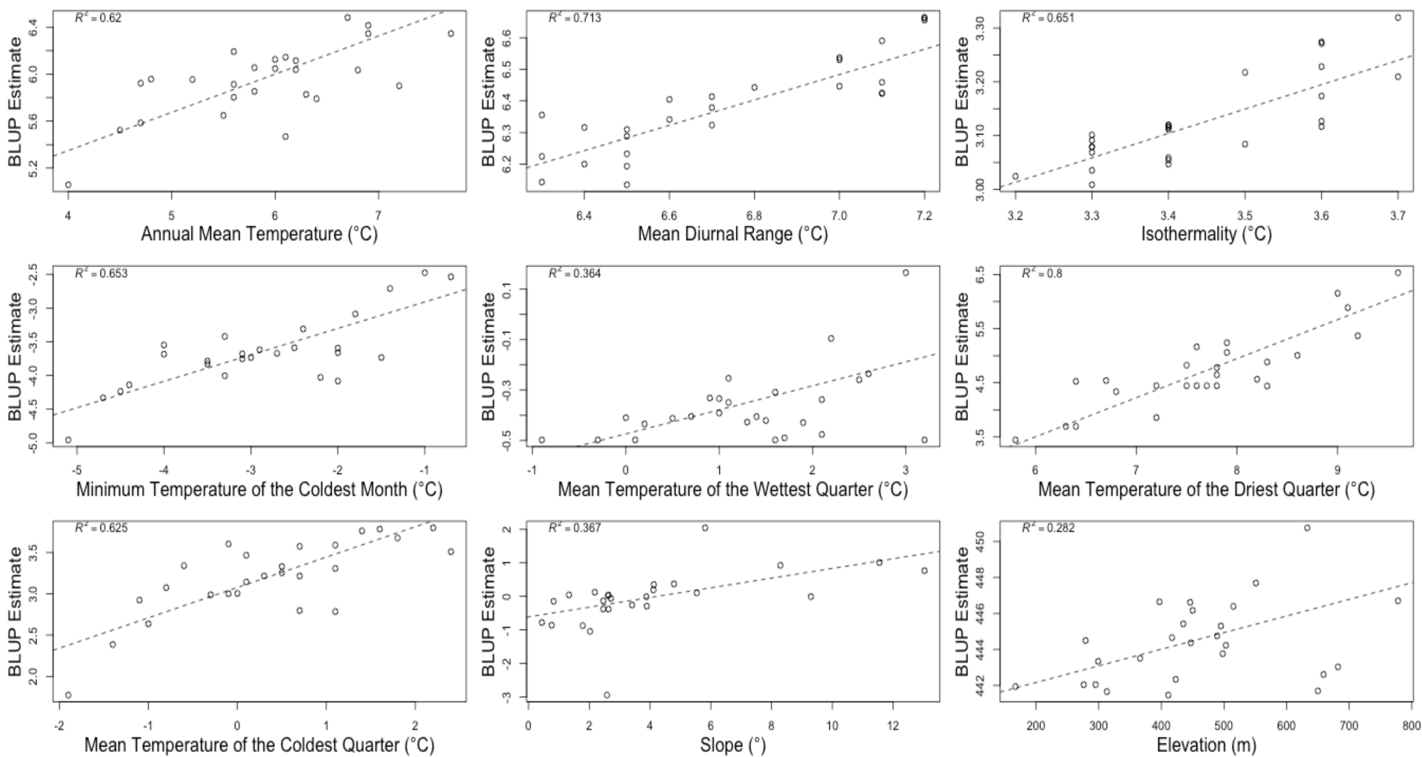


Figure S8. BLUP plots for nine environmental variables each with more than six associated loci, with dotted line denoting theoretical optimum genotype.

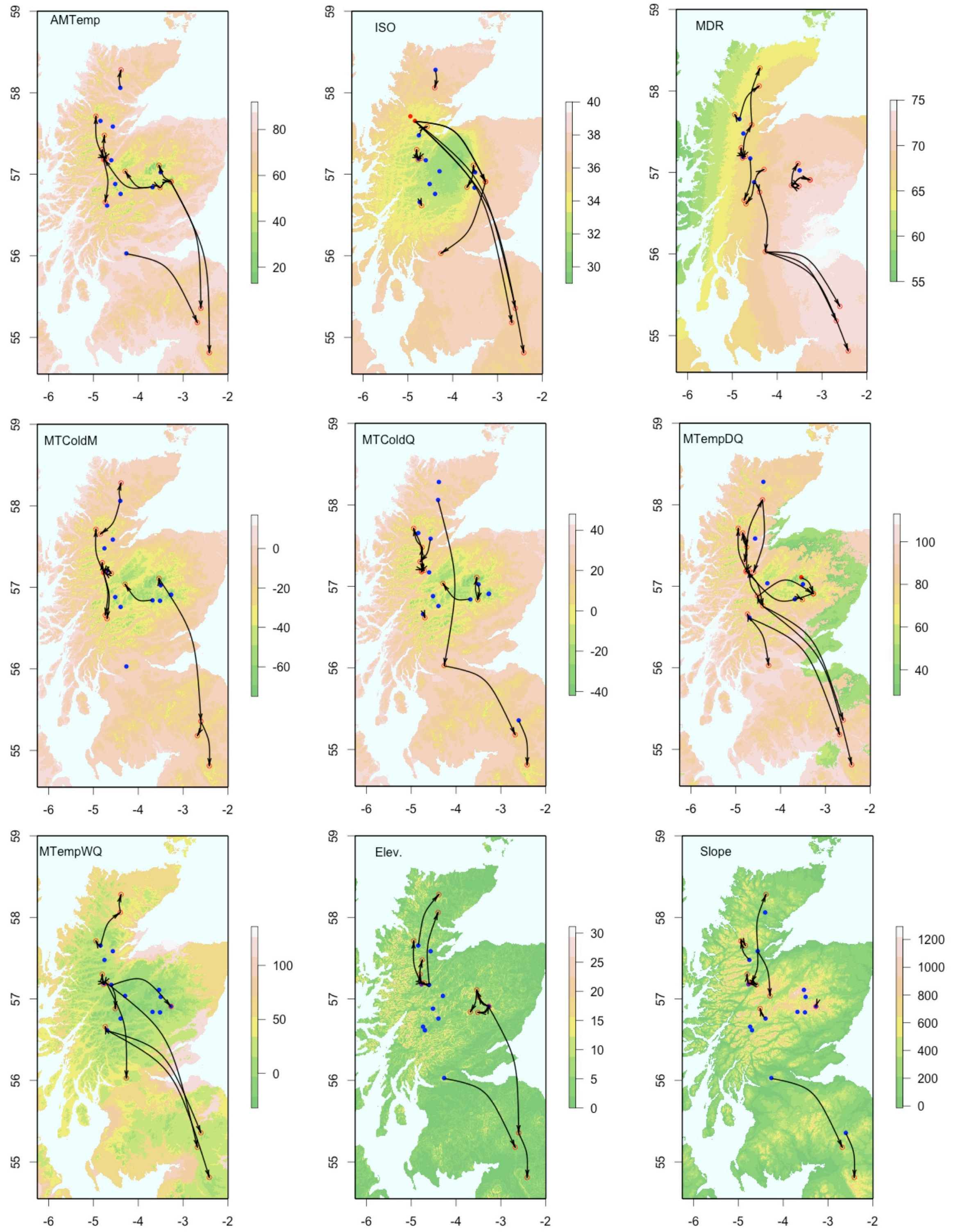


Figure S9. Assisted gene flow maps for the nine environmental variables with more than six significantly associated loci.

Table S1. Primer sequences, product lengths and general characteristics of 24 microsatellite

Origin	Genbank No.	Genome	Repeat	Forward Sequence	Reverse Sequence	Annealing Temp.	Predicted length	Product length	No. Alleles	Multiplex panel
Kulju *et al.* 2004	AF310875	Nuclear	$(GA)_6TA(GA)_6$	TGGTTGACGTGACGTTGATT	GGCCCATAGGGAAGATAAGC	64	210–222	207-251	39	Panel 1
Kulju *et al.* 2004	AF310856	Nuclear	$(AG)_4AA$	TTTCCAACGCTTTCTTGATG	TGGATAAGGAAGGGCATGTC	64	152–206	166-201	30	Panel 1
Kulju *et al.* 2004	AF310871	Nuclear	$(CA)_3(GA)_{14}$	CACCACCACAACCACCATTA	AACACCCTTTGCAACAATGA	64	93–108	173-215	32	Panel 1
Kulju *et al.* 2004	AF310851	Nuclear	$(CT)_{13}A(TC)_6$	CACACTGCTGCCTGA	TCATAAAACCCTCAAAGAAT	64	134–166	209-217	8	Panel 1
Kulju *et al.* 2004	AF310854	Nuclear	$(CT)_{12}CCTT(CT)_4$	GTTTTGGGTTTCCACTTCCA	ACTGGTAATACCTTTACCAAGCC	64	146–152	139-154	14	Panel 1
Kulju *et al.* 2004	AF310864	Nuclear	$(GT)_{18}(GA)_{14}$	GGGGATCCAGTAAGCGGTAT	CACACGAGAGATAGAGTAACGGAA	64	178–226	195-211	16	Panel 1
Kulju *et al.* 2004	AF310877	Nuclear	$(CT)_{14}$	TCTACGCTGTGACCAGTC	AGAATCCTAGCCTTTTCAAT	55	168–236	172-218	27	Panel 2
Truong *et al.* 2005	AF310847	Nuclear	$(AG)_{16}$	CGGGAAGATATGCAGTGTTT	TTGGCGGGTGAAGTAGAC	58	208–252	205-217	8	Panel 2
Truong *et al.* 2005	AF310848	Nuclear	$(CT)_9$	CTATATTGGCTCCAAGCAC	ACACCCACACTGACAGATAA	55	94–128	94-121	17	Panel 2
Truong *et al.* 2005	AY423611	Nuclear	$(TC)_{14}$	TGGCAGCACGAAAGT	TGGGAATGAGAGAACAAG	48	172–210	176-218	35	Panel 2
Kulju *et al.* 2004	AF310866	Nuclear	$(CT)_{11}$	GGCCAACAGATATAAAACGACG	TTTTAAATGCCCACCTTCCC	48	295–307	284-312	27	Panel 2
Mehlenbacher *et al.* 2010	AY423613	Nuclear	$(AG)_{17}$	CTTACCGTCCTGCCAAGGT	ACCACCACAGCCACAACC	60	235	208-242	30	Panel 2
De novo	-	Nuclear	$(CT)_5TGCTGGC(T)_3(CA)_7$	TATCAGAGACCAATGCCCAAG	CCAGGGAGTGTAGACAAAGGAA	60	282	279-294	16	Panel 3
De novo	-	Nuclear	$(GA)_{24}$	TAAATACAACCTCTCTCGCTCTAG	AGGCATTGTGGCAGAAAATC	58	330	279-334	42	Panel 3
De novo	-	Nuclear	$(TC)_{15}$	ATCTAAGGCACCCGTTCTCT	GAAAATATCTCTGCGTGCG	58	303	290-321	20	Panel 3
De novo	-	Nuclear	$(AT)_{10}$	CTTTTACGCCTTGGCTATCAA	CAGTGTCCATCTCTAGTTCAAAATG	59	167	155-197	37	Panel 3
De novo	-	Chloroplast	$(AT)_5$	GATTCACAAGTCCAATCCCAAT	TGGCTGGAAAAGTTCTATTGCT	60	235	233-235	3	Panel 3
De novo	-	Mitochondria	$(AT)_6$	TATAGTGGGGTTGAAACACGAG	GACCGTTCGCTTTCTCAGTAG	59	299	298-316	4	Panel 3
De novo	-	Nuclear	$(TC)_{17}$	TCGAAAAGGATGATGAGGATCT	ACCTGAACCAACAACGACAAC	60	243	229-245	11	Panel 4
De novo	-	Nuclear	$(AC)_{15}$	CTCATTCACTATTGTACTTGTTGGC	ATGGAAAAGAAACCGCACA	59	153	136-165	20	Panel 4
De novo	-	Chloroplast	$(AT)_5$	CATGGATGAGGTACTAGATGGTGA	AACCAACCCAAACACAAAACTC	60	383	Poor	2	Panel 4
De novo	-	Chloroplast	$(T)_{13}$	CTTTGGTGAGATCCAAGAGTTTC	TTATATGTCCACCTTGTTCCCC	60	308	Poor	3	Panel 4
De novo	-	Chloroplast	$(TA)_9$	ATTCTGCGTACCATATCCCAAA	GGCCCCTTGTAACTTCTAACAAA	60	201	Poor	2	Panel 4
De novo	-	Chloroplast	$(T)_{10}$	AGCTTTATCATCCTGGTGTTGAG	AGAAGTTTGAGCCTTTGTTTGC	60	332	Poor	3	Panel 4

Table S2. Microsatellite prior parameters used in DIY-ABC analysis.

Marker	Parameter	Distribution	Minimum	Maximum	Mean
PCR-SSR	Mean mutation rate (u)	uniform	1.00E-04	4.00E-03	0.0005
	Individual locus mutation rate	gamma	1.00E-05	1.00E-02	Mean u
	Mean coefficient (P)	uniform	1.00E-01	4.00E-01	0.22
	Individual locus coefficient	gamma	1.00E-02	9.00E-01	Mean P
RAD-SSR	Mean mutation rate (u)	uniform	1.00E-04	1.00E-03	0.0005
	Individual locus mutation rate	gamma	1.00E-05	1.00E-02	Mean u
	Mean coefficient (P)	uniform	1.00E-01	3.00E-01	0.22
	Individual locus coefficient	gamma	1.00E-02	9.00E-01	Mean P

Table S3. Proportion of *B. pubescens* introgression in *B. nana* populations across the UK.

ID	Introgression
BL	0.011
MO	0.017
BE	0.011
LH	0.015
BW	0.01
ME	0.031
GC	0.004
DE	0.005
AS	0.026
BB	0.006
PC	0.044
AV	0.013
MD	0.019
SL	0.021
MU1	0.027
MU2	0.021
LG	0.025
LL	0.023
BG	0.002
LR	0.014
RW	0.02
RB	0.02
LX	0.04
EM	0.022
SA	0.011
TD	0.014
Mean	0.018

Table S4. Adaptive deficit under current and future climate scenarios, excluding associations with altitude and slope.

ID	ADI	2045-2065				2081-2100			
	(no-alt/slope)	RCP2.6	RCP4.5	RCP6.0	RCP8.5	RCP2.6	RCP4.5	RCP6.0	RCP8.5
BL	0.216	0.272	0.311	0.262	0.375	0.280	0.236	0.254	0.242
MO	0.174	0.243	0.286	0.231	0.360	0.247	0.247	0.255	0.218
BE	0.463	0.455	0.478	0.458	0.521	0.454	0.405	0.483	0.426
LH	0.196	0.167	0.243	0.190	0.306	0.187	0.199	0.169	0.189
BW	0.113	0.135	0.205	0.104	0.218	0.127	0.089	0.134	0.103
ME	0.148	0.255	0.104	0.244	0.165	0.225	0.203	0.233	0.202
GC	0.071	0.182	0.076	0.159	0.086	0.188	0.116	0.188	0.113
DE	0.215	0.338	0.275	0.175	0.274	0.338	0.231	0.304	0.235
AS	0.259	0.278	0.220	0.274	0.223	0.294	0.267	0.300	0.305
BB	0.312	0.398	0.392	0.433	0.354	0.417	0.294	0.419	0.323
PC	0.109	0.117	0.202	0.107	0.163	0.114	0.157	0.112	0.161
AV	0.299	0.333	0.407	0.296	0.323	0.322	0.300	0.317	0.290
MD	0.230	0.422	0.374	0.453	0.344	0.415	0.214	0.436	0.226
SL	0.075	0.080	0.108	0.085	0.186	0.074	0.048	0.081	0.088
MU1	0.255	0.244	0.210	0.243	0.320	0.236	0.228	0.258	0.240
MU2	0.193	0.241	0.261	0.206	0.310	0.207	0.211	0.233	0.205
LG	0.096	0.160	0.136	0.180	0.135	0.170	0.123	0.167	0.133
LL	0.087	0.078	0.219	0.087	0.201	0.089	0.081	0.069	0.086
BG	0.250	0.254	0.185	0.251	0.124	0.237	0.240	0.238	0.232
LR	0.135	0.130	0.083	0.123	0.075	0.142	0.095	0.093	0.104
RW	0.248	0.240	0.236	0.206	0.243	0.170	0.212	0.236	0.263
RB	0.190	0.159	0.194	0.184	0.234	0.190	0.216	0.203	0.172
LX	0.339	0.523	0.538	0.497	0.387	0.535	0.327	0.523	0.350
EM	0.332	0.387	0.398	0.397	0.388	0.260	0.324	0.242	0.369
SA	0.364	0.329	0.360	0.322	0.333	0.357	0.374	0.344	0.375
TD	0.348	0.421	0.385	0.433	0.365	0.431	0.355	0.436	0.342

Table S5. Shapley values for neutral and putative adaptive loci ordered by rank, population AD and consensus ranking with Shapley adaptive contribution maximized and AD minimized.

Pop	Shapley (neutral)	Pop	Shapley (adaptive)	Pop	AD	Consensus rank
BG	0.422	SA	0.148	GC	0.045	GC
SA	0.350	BG	0.073	LG	0.064	SL
EM	0.155	BW	0.071	PC	0.081	BW
TD	0.133	LX	0.063	SL	0.085	LR
AS	0.119	TD	0.048	LR	0.097	BL
LX	0.102	MU2	0.046	LL	0.106	DE
SL	0.035	SL	0.043	ME	0.128	ME
GC	0.027	BL	0.042	LH	0.131	LH
BL	0.011	DE	0.041	BW	0.149	MO
AV	0.010	ME	0.039	MO	0.168	BG
BW	0.010	MO	0.027	RB	0.169	RB
MD	0.010	LH	0.025	DE	0.174	LX
BE	0.010	GC	0.022	BL	0.194	MU2
DE	0.009	MU1	0.020	BG	0.194	MU1
LR	0.008	LR	0.019	MU2	0.218	RW
BB	0.008	BB	0.017	RW	0.218	TD
RB	0.008	RW	0.016	AS	0.219	LL
PC	0.008	EM	0.016	MD	0.222	MD
MU2	0.008	RB	0.014	MU1	0.223	EM
LH	0.008	MD	0.008	LX	0.241	BB
RW	0.007	AV	0.007	EM	0.254	AV
LG	0.007	LL	0.007	TD	0.291	PC
MO	0.006	BE	0.006	AV	0.306	SA
MU1	0.006	PC	0.005	SA	0.321	LG
ME	0.005	LG	0.004	BB	0.366	AS
LL	0.005	AS	0.004	BE	0.479	BE

References

Aægisdóttir, H. H., Kuss, P., & Stöcklin, J. (2009). Isolated populations of a rare alpine plant show high genetic diversity and considerable population differentiation. *Annals of Botany*, *104*(7), 1313–1322.

Abebe, T. D., Naz, A. A., & Léon, J. (2015). Landscape genomics reveal signatures of local adaptation in barley (*Hordeum vulgare* L.). *Frontiers in Plant Science*, *6:813*.

Aguilar, R., Quesada, M., Ashworth, L., Herrerias-Diego, Y., & Lobo, J. (2008). Genetic consequences of habitat fragmentation in plant populations: Susceptible signals in plant traits and methodological approaches. *Molecular Ecology*, *17*(24), 5177–5188.

Aitken, S. N., & Bemmels, J. B. (2016). Time to get moving: Assisted gene flow of forest trees. *Evolutionary Applications*, *9*(1), 271–290.

Aitken, S. N., & Whitlock, M. C. (2013). Assisted Gene Flow to Facilitate Local Adaptation to Climate Change. *Annual Review of Ecology, Evolution, and Systematics*, *44*, 367–88.

Aitken, S. N., Yeaman, S., Holliday, J. a., Wang, T., & Curtis-McLane, S. (2008). Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, *1*(1), 95–111.

Alberto, F. J., Aitken, S. N., Alía, R., González-Martínez, S. C., Hänninen, H., Kremer, A., … Savolainen, O. (2013). Potential for evolutionary responses to climate change - evidence from tree populations. *Global Change Biology*, *19*(6), 1645–1661.

Alfaro, R. I., Fady, B., Vendramin, G. G., Dawson, I. K., Fleming, R. A., Saenz-Romero, C., … Loo, J. (2014). The role of forest genetic resources in responding to biotic and abiotic factors in the context of anthropogenic climate change. *Forest Ecology and Management*, *333*, 76–87.

Alsos, I. G., Eidesen, P. B., Ehrich, D., Skrede, I., Westergaard, K., Jacobsen, G. H., … Brochmann, C. (2007). Frequent long-distance plant colonization in the changing Arctic. *Science*, *316*(5831), 1606–9.

Alsos, I. G., Engelskjn, T., & Brochmann, C. (2002). Conservation Genetics and Population History of *Betula nana*, *Vaccinium uliginosum*, and *Campanulao rotundifolia* in the Arctic Archipeligo of Svalbard. *Arctic, Antarctic and Alpine Research*, *34*(4), 408–418.

Alsos, I. G., Spjelkavik, S., & Engelskjøn, T. (2003). Seed bank size and composition of *Betula nana, Vaccinium uliginosum*, and *Campanula rotundifolia* habitats in Svalbard and northern Norway. *Canadian Journal of Botany*, *81*(3), 220–231.

Anamthawat-Jónsson, K., & Thórsson, A. T. (2003). Natural hybridisation in birch: Triploid hybrids between *Betula nana* and *B. pubescens*. *Plant Cell, Tissue and Organ Culture*, *75*(2), 99–107.

Anamthawat-Jónsslon, K., & Tomasson, T. (1999). High frequency of triploid birch hybrid by Betula nana seed parent. *Hereditas*, *130*(2), 191–193.

Anderson, J. T. (2016). Plant fitness in a rapidly changing world. *New Phytologist*, *210*(1), 81–87.

Anderson, J. T., Willis, J. H., & Mitchell-olds, T. (2011). Evolutionary genetics of plant

adaptation. *Trends in Genetics*. *27*(7), 258–266.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*, 81–92.

Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., & Luikart, G. (2008). LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, *9*:323.

Arnold, M. L., & Martin, N. H. (2009). Adaptation by introgression. *Journal of Biology*, *8(9)*.

Ashburner, K., & McAllister, H. A. (2013). The genus *Betula*: a taxonomic revision of birches. (M. Rix, Ed.). Royal Botanic Gardens, Kew.

Asmyhr, M. G., Linke, S., Hose, G., & Nipperess, D. A. (2014). Systematic conservation planning for groundwater ecosystems using phylogenetic diversity. *PLoS ONE*, *9*(12), 1–15.

Aston, D. (1984). *Betula nana* L., a note on its status in the United Kingdom. *Proceedings of the Royal Society of Edinburgh Section B: Biological Sciences*, *85*(1-2), 43–47.

Atkinson, M. D. (1992). *Betula Pendula Roth (B . Verrucosa Ehrh.)* and *B. Pubescens Ehrh. Journal of Ecology*. 80:4, 837-870.

Bacles, C. F. E., Burczyk, J., Lowe, A. J., & Ennos, R. A. (2005). Historical and Contemporary Mating Patterns in Remnant Populations of the Forest Tree *Fraxinus Excelsior* L . *Evolution*, *59*(5), 979–990.

Bacles, C. F. E., & Jump, A. S. (2011). Taking a tree's perspective on forest fragmentation genetics. *Trends in Plant Science*. 16(1). 13-18.

Bacles, C. F. E., Lowe, A. J., & Ennos, R. a. (2006). Effective seed dispersal across a fragmented landscape. *Science*. *311*(5761), 628.

Bacles, C. F. E., Lowe, A. J., & Ennos, R. A. (2004). Genetic effects of chronic habitat fragmentation on tree species: The case of Sorbus aucuparia in a deforested Scottish landscape. *Molecular Ecology*, *13*(3), 573–584.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., … Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS One*, *3*(10), e3376.

Bayfield, N. G. (1979). Recovery of four montane heath communities on Cairngorm, Scotland, from disturbance by trampling. *Biological Conservation*, *15*(3), 165–179.

Beatty, G. E., Philipp, M., & Provan, J. (2010). Unidirectional hybridization at a species' range boundary: Implications for habitat tracking. *Diversity and Distributions*, *16*(1), 1–9.

Beatty, G. E., & Provan, J. (2011). Comparative phylogeography of two related plant species with overlapping ranges in Europe, and the potential effects of climate change on their intraspecific genetic diversity. *BMC Evolutionary Biology*, *11*(1), 29.

Beaumont, M. A., & Nichols, R. A. (1996). Evaluating Loci for Use in the Genetic Analysis of Population Structure. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *263*(1377), 1619–1626.

Becker, M., Gruenheit, N., Steel, M., Voelckel, C., Deusch, O., Heenan, P. B., … Lockhart, P. J.

(2013). Hybridization may facilitate in situ survival of endemic species through periods of climate change. *Nature Climate Change*, *3*(12), 1039–1043.

Benestan, L., Ferchaud, A.L., Hohenlohe, P., Garner, B. A., Naylor, G. J. P., Baums, I., … Luikart, G. (2016). Conservation genomics of natural and managed populations: building a conceptual and practical framework. *Molecular Ecology*, 25, 2967–2977.

Bennett, K.D., Tzedakis, P.D., & Willis, K. J. (1991). Quaternary refugia of north European trees. *Journal of Biogeography.* 18(1), 103-115.

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*(3-4), 129–151.

Bertorelle, G., Benazzo, A., & Mona, S. (2010). ABC as a flexible framework to estimate demography over space and time: Some cons, many pros. *Molecular Ecology*, *19*(13), 2609–2625.

Bhagwat, S. A., & Willis, K. J. (2008). Species persistence in northerly glacial refugia of Europe: A matter of chance or biogeographical traits? *Journal of Biogeography*, *35*(3), 464–482.

Bierne, N., Roze, D., & Welch, J. J. (2013). Pervasive selection or is it …? why are FST outliers sometimes so frequent ? *Molecular Ecology*, *22*, 2061–2064.

Bingham, R. a, & Ranker, T. a. (2011). Genetic Diversity in Alpine and Foothill Populations of *Campanula rotundifolia. International Journal of Plant Sciences*, *161*(3), 403-411.

Birks, H. H. (1994). Plant macrofossils and the nunatak theory of pre-glacial survival. *Dissertationes Botanicae*, *234*, 129–143.

Birks, H. J. B. (1968). The identification of Betula nana pollen. *New Phytologist*, *67*, 309–314.

Birks, H. J. B., & Willis, K. J. (2008). Alpines, trees, and refugia in Europe. *Plant Ecology & Diversity*, *1*(2), 147–160.

Bjorkman, A. D., Vellend, M., Frei, E. R., & Henry, G. H. R. (2016). Climate adaptation is not enough: warming does not facilitate success of southern tundra plant populations in the high Arctic. *Global Change Biology*, 1–12.

Blair, L. M., Granka, J. M., & Feldman, M. W. (2014). On the stability of the Bayenv method in assessing human SNP-environment associations. *Human Genomics*, *8*(1).

Blanquart, F., Kaltz, O., Nuismer, S. L., & Gandon, S. (2013). A practical guide to measuring local adaptation. *Ecology Letters*, *16*(9), 1195–1205.

Bohlen, J., Slechtova, V., Slechta, V., Slechtova, V., Sember, A., & Rab, P. (2016). A Ploidy Difference Represents an Impassable Barrier for Hybridisation in Animals. Is There an Exception among Botiid Loaches (Teleostei: *Botiidae*)? *PLoS ONE*, *11*(7).

Bolliger, J., Lander, T., & Balkenhol, N. (2014). Landscape genetics since 2003: Status, challenges and future directions. *Landscape Ecology*, *29*(3), 361–366.

Boshier, D., Broadhurst, L., Cornelius, J., Gallo, L., Koskela, J., Loo, J., … St Clair, B. (2015). Is local best? Examining the evidence for local adaptation in trees and its scale. *Environmental Evidence*, *4*(20).

Boyd, I. L., Freer-Smith, P. H., Gilligan, C. A, & Godfray, H. C. J. (2013). The consequence of tree pests and diseases for ecosystem services. *Science*, *342*, 823-829.

Bradbury, I. R., Hamilton, L. C., Dempson, B., Robertson, M. J., Bourret, V., Bernatchez, L., & Verspoor, E. (2015). Transatlantic secondary contact in Atlantic Salmon, comparing microsatellites, a single nucleotide polymorphism array and restriction-site associated DNA sequencing for the resolution of complex spatial structure. *Molecular Ecology*, *24*(20), 5130–5144.

Breed, M. F., Stead, M. G., Ottewell, K. M., Gardner, M. G., & Lowe, A. J. (2013). Which provenance and where? Seed sourcing strategies for revegetation in a changing environment. *Conservation Genetics*, *14*(1), 1–10.

Brennan, A. C., Woodward, G., Seehausen, O., Munoz-Fuentes, V., Moritz, C., Guelmami, A., … Edelaar, P. (2014). Hybridization due to changing species distributions: Adding problems or solutions to conservation of biodiversity during global change? *Evolutionary Ecology Research*, *16*(6), 475–491.

Brewer, S., Cheddadi, R., Beaulieu, J. L. De, & Reille, M. (2002). The spread of deciduous *Quercus* throughout Europe since the last glacial period. *Forest Ecology and Management*. *156,* 27-48.

Brooks, T. M., Mittermeier, R. A., da Fonseca, G. A. B., Gerlach, J., Hoffmann, M., Lamoreux, J. F., … Rodrigues, A. S. L. (2006). Global Biodiversity Conservation Priorities. *Science*, *313*, 58–61.

Bryant, D., & Moulton, V. (2004). Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. *Molecular Biology and Evolution*, *21*(2), 255–265.

Buggs, R. J. A. (2007). Empirical study of hybrid zone movement. *Heredity*, *99*(3), 301–12.

Bunce, R. G. H., Wood, C. M., Smart, S. M., Oakley, R., Browning, G., Daniels, M. J., … Cresswell, J. (2014). The Landscape Ecological Impacts of Afforestation in the British Uplands and Some Initiatives to Restore Native Woodland Cover. *Journal of Landscape Ecology*, *7*(2), 5–24.

Buza, L., Young, A., & Thrall, P. (2000). Genetic erosion, inbreeding and reduced fitness in fragmented populations of the endangered tetraploid pea *Swainsona recta*. *Biological Conservation*, *93*(2), 177–186.

Caseldine, C. (2001). Changes in *Betula* in the Holocene record from Iceland - A palaeoclimatic record or evidence for early Holocene hybridisation? *Review of Palaeobotany and Palynology*, *117*(1-3), 139–152.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *Genes, Genomes, Genetics. 1*(3), 171–182.

Cavender-Bares, J., González-Rodríguez, A., Eaton, D. A. R., Hipp, A. A. L., Beulke, A., & Manos, P. S. (2015). Phylogeny and biogeography of the american live oaks (*Quercus* subsection *Virentes*): A genomic and population genetics approach. *Molecular Ecology*, *24*(14), 3668–3687.

Cheddadi, R., Vendramin, G. G., Litt, T., François, L., Kageyama, M., Lorentz, S., … Lunt, D. (2006). Imprints of glacial refugia in the modern genetic diversity of *Pinus sylvestris*. *Global Ecology and Biogeography*, *15*(3), 271–282.

Chen, J., Gyllenstrand, N., Zaina, G., Morgante, M., Bousquet, J., Eckert, A., … Lascoux, M. (2012). Disentangling the Roles of History and Local Selection in Shaping Clinal Variation of Allele Frequencies and Gene Expression in Norway Spruce (*Picea abies*), *191*, 1–17.

Chevin, L. M., & Lande, R. (2011). Adaptation to marginal habitats by evolution of increased phenotypic plasticity. *Journal of Evolutionary Biology*, *24*(7), 1462–1476.

Christie, K. S., Ruess, R. W., Lindberg, M. S., & Mulder, C. P. (2014). Herbivores influence the growth, reproduction, and morphology of a widespread arctic willow. *PLoS ONE*, *9*(7), 1–9.

Clark, J. S., Fastie, C., Hurtt, G., Jackson, S. T., Johnson, C., King, G. A., … others. (1998). Reid's Paradox of Rapid Plant Migration Dispersal theory and interpretation of paleoecological records. *BioScience*, *48*(1), 13–24.

Clark, P., Dyke, A., Shakun, J., Carlson, A., Clark, J., Wohlfarth, B., … McCabe, M. (2009). The Last Glacial Maximum. *Science*, *325*(5941), 710–714.

Clark, L. V., Ryan Stewart, J., Nishiwaki, A., Toma, Y., Kjeldsen, J. B., Jorgensen, U., … Sacks, E. J. (2015). Genetic structure of *Miscanthus sinensis* and *Miscanthus sacchariflorus* in Japan indicates a gradient of bidirectional but asymmetric introgression. *Journal of Experimental Botany*, *66*(14), 4213–4225.

Cleland, E.E., Allen, J.M., Crimmins, T.M., Dunne, J.A., Pau, S., Travers, S.E., … Wolkovich. (20012). Phenological tracking enables positive species responses to climate change. *Ecology*, *93*(8), 1765–1771.

Coates, B. S., Sumerford, D. V., Miller, N. J., Kim, K. S., Sappington, T. W., Siegfried, B. D., & Lewis, L. C. (2009). Comparative performance of single nucleotide polymorphism and microsatellite markers for population genetic analysis. *Journal of Heredity*, *100*(5), 556–564.

Collen, B., Turvey, S. T., Waterman, C., Meredith, H. M. R., Kuhn, T. S., Baillie, J. E. M., & Isaac, N. J. B. (2011). Investing in evolutionary history: implementing a phylogenetic approach for mammal conservation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *366*(1578), 2611–2622.

Collin, E. (2003). EUFORGEN Technical guidelines for genetic conservation and use of European white elm, *Ulmus laevis*. ISBN 92-9043-603-4.

Comes, H. P., & Kadereit, J. W. (1998). The effect of quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science*, *3*(11), 432–438.

Conesa, A., & Gotz, S. (2007). Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. *International Journal of Plant Genomics*, *12*.

Corlett, R. T., & Westcott, D. A. (2013). Will plant movements keep up with climate change? *Trends in Ecology and Evolution*, *28*(8), 482–488.

Cornille, A., Giraud, T., Bellard, C., Tellier, A., Le Cam, B., Smulders, M. J. M., … Gladieux, P. (2013). Postglacial recolonization history of the European crabapple (*Malus sylvestris* Mill.), a wild contributor to the domesticated apple. *Molecular Ecology*, *22*(8), 2249–

2263.

Cornuet, J. M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., … Estoup, A. (2014). DIYABC v2.0: A software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, *30*(8), 1187–1189.

Crispo, E., Moore, J. S., Lee-Yaw, J. A., Gray, S. M., & Haller, B. C. (2011). Broken barriers: Human-induced changes to gene flow and introgression in animals: An examination of the ways in which humans increase genetic exchange among populations and species and the consequences for biodiversity. *BioEssays*, *33*(7), 508–518.

Dąbrowska, G., Dzialuk, A., Burnicka, O., Ejankowski, W., Gugnacka-Fiedor, W., & Goc, A. (2006). Genetic diversity of postglacial relict shrub *Betula nana* revealed by RAPD analysis. *Dendrobiology*, *55*, 19–23.

Davies, S. J., Cavers, S., Finegan, B., White, A., Breed, M. F., & Lowe, A. J. (2013). Pollen flow in fragmented landscapes maintains genetic diversity following stand-replacing disturbance in a neotropical pioneer tree, *Vochysia ferruginea* Mart. *Heredity*, *115*, 1–5.

Davis, M. B., & Shaw, R. G. (2001). Range shifts and adaptive responses to Quaternary climate change. *Science*, *292*, 673–679.

Day, K. R., & Leather, S. R. (1997). Threats to forestry by insect pests in Europe. In A.D. Watt, N.E. Stork, M.D. Hunter (Eds.), *Forests and Insects*, London, Chapman & Hall.

de Bello, F., Lavorel, S., Lavergne, S., Albert, C. H., Boulangeat, I., Mazel, F., & Thuiller, W. (2013). Hierarchical effects of environmental filters on the functional structure of plant communities: A case study in the French Alps. *Ecography*, *36*(3), 393–402.

de Lafontaine, G., Amasifuen Guerra, C. A., Ducousso, A., Sánchez-Goñi, M.-F., & Petit, R. J. (2014). Beyond skepticism: uncovering cryptic refugia using multiple lines of evidence. *New Phytologist*, *204*(3), 450–454.

De Mita, S., Thuillet, A.-C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., & Vigouroux, Y. (2013). Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, *22*(5), 1383–1399.

Deblauwe, V., Droissart, V., Bose, R., Sonké, B., Blach-Overgaard, A., Svenning, J. C., … Couvreur, T. L. P. (2016). Remotely sensed temperature and precipitation data improve species distribution modelling in the tropics. *Global Ecology and Biogeography*, *25*(4), 443–454.

Dillon, S., McEvoy, R., Baldwin, D. S., Rees, G. N., Parsons, Y., & Southerton, S. (2014). Characterisation of adaptive genetic diversity in environmentally contrasted populations of Eucalyptus camaldulensis Dehnh. (river red gum). *PLoS ONE*, *9*(8).

Diniz-Filho, J. A. F., Melo, D. B., de Oliveira, G., Collevatti, R. G., Soares, T. N., Nabout, J. C., … Telles, M. P. de C. (2012). Planning for optimal conservation of geographical genetic variability within species. *Conservation Genetics*, *13*(4), 1085–1093.

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, *14*, 927–930.

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014).

NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, *14*(1), 209–214.

Dodgshon, R. A., & Olsson, G. A. (2006). Heather moorland in the Scottish Highlands: The history of a cultural landscape, 1600-1880. *Journal of Historical Geography*, *32*(1), 21–37.

Dongen, S. Van, Backeljau, T., Matthysen, E., & Dhondt, A. A. (1994). Effects of forest fragmentation on the population structure of the Study areas. Acta Ecologica. *15*(2), 193-206.

Douda, J., Doudová, J., Drašnarová, A., Kuneš, P., Hadincová, V., Krak, K., … Mandák, B. (2014). Migration patterns of subgenus Alnus in Europe since the last glacial maximum: a systematic review. *PloS One*, *9*(2), e88709.

Doyle, J., Doyle, J. L. (1987). Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem Bull*, *19*(11), 11–15.

Drzymulska, D. (2014). Postglacial occurrence and decline of *Betula nana* L. (dwarf birch) in northeastern Poland. *Estonian Journal of Earth Sciences*, *63(2),* 76-87.

Dubreuil, M., Riba, M., González-Martínez, S. C., Vendramin, G. G., Sebastiani, F., & Mayol, M. (2010). Genetic effects of chronic habitat fragmentation revisited: Strong genetic structure in a temperate tree, *Taxus baccata* (Taxaceae), with great dispersal capability. *American Journal of Botany*, *97*(2), 303–310.

Dufresne, F., Stift, M., Vergilino, R., & Mable, B. K. (2014). Recent progress and challenges in population genetics of polyploid organisms: An overview of current state-of-the-art molecular and statistical tools. *Molecular Ecology*, *23*(1), 40–69.

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361.

Eckert, A. J., Wegrzyn, J. L., Pande, B., Jermstad, K. D., Lee, J. M., Liechty, J. D., … Neale, D. B. (2009). Multilocus patterns of nucleotide diversity and divergence reveal positive selection at candidate genes related to cold hardiness in coastal Douglas fir (*Pseudotsuga menziesii var. menziesii*). *Genetics*, *183*(1), 289–298.

Eckert, C. G., Samis, K. E., & Lougheed, S. C. (2008). Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology*, *17*(5), 1170–1188.

Ehrlich, P. R. (1996). Conservation in temperate forests: What do we need to know and do? *Forest Ecology and Management*, *85*(1-3), 9–19.

Eidesen, P. B., Alsos, I. G., & Brochmann, C. (2015). Comparative analyses of plastid and AFLP data suggest different colonization history and asymmetric hybridization between *Betula pubescens* and *B. nana*. *Molecular Ecology*, *24*(15), 3993–4009.

Eidesen, P. B., Ehrich, D., Bakkestuen, V., Alsos, I. G., Gilg, O., Taberlet, P., & Brochmann, C. (2013). Genetic roadmap of the Arctic: Plant dispersal highways, traffic barriers and capitals of diversity. *New Phytologist*, *200*(3), 898–910.

Eizaguirre, C., & Baltazar-Soares, M. (2014). Evolutionary conservation-evaluating the adaptive potential of species. *Evolutionary Applications*, *7*(9), 963–967.

Ejankowski, W. (2008). Effect of waterlogging on regeneration in the dwarf birch (*Betula nana*). *Biologia*, *63*(5), 670–676.

Ejankowski, W. (2010). Demographic variation of dwarf birch (*Betula nana*) in communities dominated by *Ledum palustre* and *Vaccinium uliginosum*. *Biologia*, *65*, 248–253.

Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, *1*(4), 330–342.

Elith, J., & Leathwick, J. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution*, *40*, 677–697.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, *17*(1), 43–57.

Elkington, T. (1968). Introgressive hybridization between *Betula nana* L. and *B. pubescens* Ehrh. in North-West Iceland. *New Phytologist*, *67*, 109–118.

Ellegren, H. (2009). A selection model of molecular evolution incorporating the effective population size. *Evolution*, *63*(2), 301–305.

Ellstrand, N. C., & Elam, D. R. (1993). Population genetic consequences of small population size: Implications for plant conservation. *Annual Review of Ecology Evolution and Systematics*, *24*, 217–242.

Elsen, P. R., & Tingley, M. W. (2015). Global mountain topography and the fate of montane species under climate change. *Nature Climate Change*, *5*(8), 772–776.

Endelman, J. B. (2011). Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome Journal*, *4*(3), 250–255.

Engler, R., Randin, C. F., Thuiller, W., Dullinger, S., Zimmermann, N. E., Araújo, M. B., … Guisan, A. (2011). 21st century climate change threatens mountain flora unequally across Europe. *Global Change Biology*, *17*(7), 2330–2341.

Ennos, R. A., Worrell, R., & Malcolm, D. C. (1998). The genetic management of native species in Scotland. *Forestry*, *71*(1), 1–23.

Epps, C. W., & Keyghobadi, N. (2015). Landscape genetics in a changing world: disentangling historical and contemporary influences and inferring change. *Molecular Ecology*, *24*, 6021–6040.

Estreguil, C., Caudullo, G., de Rigo, D., & San-Miguel-Ayanz, J. (2013). Forest Landscape in Europe: Pattern, Fragmentation and Connectivity. *EUR – Scientific and Technical Research*, *25717*(JRC 77295). ISBN: 978-92-79-28118-1.

Evangelista, A., Frate, L., Carranza, M. L., Attorre, F., Pelino, G., & Stanisci, A. (2016). Changes in composition, ecology and structure of high-mountain vegetation: A re-visitation study over 42 years. *AoB PLANTS*, *8*, 1–11.

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

Excoffier, L., & Ray, N. (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology and Evolution*, *23*(7), 347–351.

Fahrig, L. (2003). Effects of Habitat Fragmentation on Biodiversity. *Annual Review of Ecology,*

Evolution, and Systematics, *34*(2003), 487–515.

Falk, D.A., & Holsinger, K. E. (Eds.) (1991). *Genetics and Conservation of Rare Plants*. Oxford University Press, New York.

FAO. (2014). *State of the World's Forests: Food and Agriculture orginization of the United Nations*. ISBN: 978-92-5-108269-0. (Available at http://www.fao.org/3/a-i3710e.pdf).

FAO. (2015). State of Europe's Forests. (Available at http://foresteurope.org/state-europes-forests-2015-report/)

FCS. (2004). Seed Sources for Planting Native Trees and Shrubs in Scotland. ISBN: 0-85538-712-2. (http://scotland.forestry.gov.uk/images/corporate/pdf/seedsourcefcfc151.pdf).

Feurdean, A., Bhagwat, S. a, Willis, K. J., Birks, H. J. B., Lischke, H., & Hickler, T. (2013). Tree migration-rates: narrowing the gap between inferred post-glacial rates and projected rates. *PloS One*, *8*(8), e71797.

Finger, A., Kettle, C. J., Kaiser-Bunbury, C. N., Valentin, T., Mougal, J., & Ghazoul, J. (2012). Forest fragmentation genetics in a formerly widespread island endemic tree: *Vateriopsis seychellarum* (Dipterocarpaceae). *Molecular Ecology*, *21*(10), 2369–2382.

Fisher, M. C., Henk, D. a, Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L., & Gurr, S. J. (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature*, *484*(7393), 186–94.

Fitzpatrick, M. C., & Keller, S. R. (2015). Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, *18*(1), 1–16.

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, *180*(2), 977–993.

Forester, B. R., Jones, M. R., Joost, S., Landguth, E. L., & Lasky, J. R. (2015). Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*, 104–120.

Forestry Commission. (2011). *The UK Forestry Standard*. *Forestry Commission*. ISBN: 978-0-85538-830-0.

Frankham, R. (2015). Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. *Molecular Ecology*, *24*(11), 2610-2618.

Frichot, E., Schoville, S., de Villemereuil, P., Gaggiotti, O. E., & François, O. (2015). Detecting adaptive evolution based on association with ecological gradients: Orientation matters! *Heredity*, (115), 22–28.

Fritz, R. S., Hochwender, C. G., Albrectsen, B. R., & Czesak, M. E. (2006). Fitness and genetic architecture of parent and hybrid willows in common gardens. *Evolution*, *60*(6), 1215–1227.

Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdelhué, C., Pudlo, P., … Estoup, A. (2013). The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology*, *22*(11), 3165–3178.

Geml, J., Kauff, F., Brochmann, C., & Taylor, D. L. (2010). Surviving climate changes: high

genetic diversity and transoceanic gene flow in two arctic–alpine lichens, *Flavocetraria cucullata* and *F. nivalis* (Parmeliaceae, Ascomycota). *Journal of Biogeography*, *37*(8), 1529–1542.

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., … Zhang, M. (2011). The community climate system model version 4. *Journal of Climate*, *24*(19), 4973–4991.

Gentili, R., Bacchetta, G., Fenu, G., Cogoni, D., Abeli, T., Rossi, G., … Citterio S. (2015). From cold to warm-stage refugia for boreo-alpine plants in southern European and Mediterranean mountains: the last chance to survive or an opportunity for speciation? *Biodiversity*, *16*(4), 247–261.

Geraldes, A., Farzaneh, N., Grassa, C. J., Mckown, A. D., Guy, R. D., Mansfield, S. D., … Cronk, Q. C. B. (2014). Landscape genomics of populus trichocarpa: The role of hybridization, limited gene flow, and natural selection in shaping patterns of population structure. *Evolution*, *68*(11), 3260–3280.

Gibson, A. L., Espeland, E. K., Wagner, V., & Nelson, C. R. (2016). Can local adaptation research in plants inform selection of native plant materials? An analysis of experimental methodologies. *Evolutionary Applications*, *9*(10) 1219-1228.

Gilbert, D., & Di Cosmo, L. (2009). Towards restoration of treeline woodland and montane scrub. *Botanical Journal of Scotland*, *55*(1), 177–187.

Glover, B. J., & Abbott, R. J. (1995). Low Genetic Diversity in the Scottish Endemic *Primula-Scotica* Hook. *New Phytologist*, *129*(1), 147–153.

Godwin, H. (1957). The History of the British Flora. *Journal of Ecology*, *45*(1), 285–287.

Godwin, H. (1975). History of the Natural Forests of Britain: Establishment, Dominance and Destruction. *Proceedings of the Royal Society B-Biological Sciences*, *271*(911).

Goldman, N., & Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution*, *11*(5), 725–736.

Gomez, J. M., Gonzalez-Megias, A., Lorite, J., Abdelaziz, M., & Perfectti, F. (2015). The silent extinction: climate change and the potential hybridization-mediated extinction of endemic high-mountain plants. *Biodiversity and Conservation*, *24*(8), 1843–1857.

González-Irusta, J., González-Porto, M., Sarralde, R., Arrese, B., Almón, B., & Martín-Sosa, P. (2015). Comparing species distribution models: a case study of four deep sea urchin species. *Hydrobiologica*, *745*(1), 43–57.

Gottfried, M., Pauli, H., Futschik, A., Akhalkatsi, M., Barančok, P., Alonso, B., … Grabherr, G. (2012). Continent-wide response of mountain vegetation to climate change. *Nature Climate Change*, *2*, 111-115.

Goudet, J. (2005). HIERFSTAT , a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *2*(5), 184–186.

Gough, L., Bass, H., & McLaren, J. R. (2015). Effects of Increased Soil Nutrients on Seed Rain: A Role for Seed Dispersal in the Greening of the Arctic? *Arctic, Antarctic, and Alpine Research*, *47*(1), 27–34.

Gramlich, S., & Horandl, E. (2016). Fitness of natural willow hybrids in a pioneer mosaic hybrid zone. *Ecology and Evolution*, *6*(21), 7645–7655.

Grant Hokit, D., Ascunce, M., Ernst, J., Branch, L. C., & Clark, A. M. (2010). Ecological metrics predict connectivity better than geographic distance. *Conservation Genetics*, *11*(1), 149–159.

Gratani, L. (2014). Plant Phenotypic Plasticity in Response to Environmental Factors. *Advances in Botany*. *2014*. 1-17.

Griffiths, A. J. F., Miller, J. H., & Suzuki, D. T. (2000). *An Introduction to Genetic Analysis*. (W. H. Freeman, Ed.) (7th edition). New York.

Groot, W. J. D., Thomas, P. A., & Wein, R. W. (1997). *Betula nana* L. and *Betula glandulosa* Michx. *Journal of Ecology*, *85*, 241–264.

Guisan, A., & Theurillat, J.-P. (2001). Assessing alpine plant vulnerability to climate change: a modeling perspective. *Integrated Assessment*, *1*(1), 307–320.

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, *8*(9), 993–1009.

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., … Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, *16*(12), 1424–1435.

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, *135*(2-3), 147–186.

Günther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics*, *195*(1), 205–220.

Haake, C.J., Kashiwada, A., & Su, F. E. (2008). The Shapley value of phylogenetic trees. *Journal of Mathematical Biology*, *56*(4), 479–497.

Hagman, M. (1971). On self- and cross-incompatibility shown by *Betula verrucosa* Ehrh. and *Betula pubescens* Ehrh. *Communicationes Instituti Forestalis Fenniae*, *73*, 125-134.

Hallatschek, O., Hersen, P., Ramanathan, S., & Nelson, D. R. (2007). Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(50), 19926–30.

Hällfors, M. H., Liao, J., Dzurisin, J. D. K., Grundel, R., Hyvärinen, M., Towle, K., … Hellmann, J. J. (2015). Addressing potential local adaptation in species distribution models: implications for conservation under climate change. *Ecological Applications*, *26*(4), 1154-1169.

Hamilton, J. A., & Miller, J. M. (2015). Adaptive introgression: a resource for management and genetic conservation in a changing climate. *Conservation Biology*, *30*(1), 33-41.

Hamlin, J. A. P., & Arnold, M. L. (2014). Determining population structure and hybridization for two iris species. *Ecology and Evolution*, *4*(6), 743–755.

Hampe, A., & Petit, R. J. (2005). Conserving biodiversity under climate change: The rear edge matters. *Ecology Letters*, *8*(5), 461–467.

Handel, S. N. (1985). The Intrusion of Clonal Growth Patterns on Plant Breeding Systems. *The American Naturalist*, *125*(3), 367–384.

Hanhimäki, S., Senn, J., & Haukioja, E. (1994). Performance of insect herbivores on hybridizing trees : the case of the subarctic birches. *Journal of Animal Ecology*, *63*(1), 163–175.

Harrisson, K. A., Pavlova, A., Telonis-Scott, M., & Sunnucks, P. (2014). Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications*, *7*(9), 1008–1025.

Havens, K., Vitt, P., Still, S., Kramer, A. T., Fant, J. B., & Schatz, K. (2015). Seed Sourcing for Restoration in an Era of Climate Change. *Natural Areas Journal*, *35*(1), 122–133.

Hecht, B. C., Matala, A. P., Hess, J. E., & Narum, S. R. (2015). Environmental adaptation in Chinook salmon (Oncorhynchus tshawytscha) throughout their North American range. *Molecular Ecology*, *24*(22), 5573–5595.

Henry, G. H. R., & Molau, U. (1997). Tundra plants and climate change: the International Tundra Experiment (ITEX). *Global Change Biology*, *3(1)*, 1–9.

Heuertz, M., Hausman, J.F., Hardy, O. J., Vendramin, G. G., Frascaria-Lacoste, N., & Vekemans, X. (2004). Nuclear microsatellites reveal contrasting patterns of genetic structure between western and southeastern European populations of the common ash (*Fraxinus excelsior* L.). *Evolution*, *58*(5), 976–988.

Hewitt, G. (1999). Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, *68*(1-2), 87–112.

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

Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence. *Biological Journal of the Linnean Society*. *58*, 247–276.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978.

Hijmans, R. J., & Etten, J. (2012). raster: Geographic analysis and modeling with raster data. Available online at: http://cran.r-project.org/package=raster

Hijmans, R. J., Philips, S., Leathwick, J., & Elith, J. (2011). Package "dismo." Available online at: http://cran.r-project.org/web/packages/dismo/index.html.

Hodel, R. G. J., Segovia-Salcedo, M. C., Landis, J. B., Crowl, A. A., Sun, M., Liu, X., … Soltis, P. S. (2016). The Report of My Death was an Exaggeration: A Review for Researchers Using Microsatellites in the 21st Century. *Applications in Plant Sciences*, *4*(6), 1600025.

Hodgins, K. A., & Moore, J. L. (2016). Adapting to a warming world: Ecological restoration, climate change, and genomics. *American Journal of Botany*, *103*(4), 590–592.

Hoersch, B., Braun, G., & Schmidt, U. (2002). Relation between landform and vegetation in alpine regions of Wallis, Switzerland. A multiscale remote sensing and GIS approach. *Computers, Environment and Urban Systems*, *26*(2-3), 113–139.

Hoffmann, A. A., Sgrò, C., & M. (2011). Climate change and evolutionary adaptation. *Nature*, *470*(7335), 479–485.

Hohenlohe, P. a, Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, *6*(2), e1000862.

Holl, K., & Smith, M. (2007). Scottish upland forests: History lessons for the future. *Forest Ecology and Management*, *249*(1-2), 45–53.

Holm, S.O. (1994). Reproductive patterns of *Betula pendula* and *B. pubescens* coll. along a regional altitudinal gradient in northern Sweden. *Ecography*, *17*(1), 60–72.

Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, *9*(5), 1322–1332.

Huck, S., Büdel, B., Kadereit, J. W., & Printzen, C. (2009). Range-wide phylogeography of the European temperate-montane herbaceous plant *Meum athamanticum* Jacq.: evidence for periglacial persistence. *Journal of Biogeography*, *36*(8), 1588–1599.

Hughes, L. (2000). Biological consequences of global warming: is the signal already apparent? *Trends in Ecology & Evolution*, *15*(2), 56–61.

Huntley, B., & Birks, H. J. B. (1983). An Atlas of Past and Present Pollen Maps for Europe, 0-13,000 Years Ago. Cambridge University Press.

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*(2), 254–267.

Hutchinson, T. C. (1966). The occurence of living and sub-fossil remains of *Betula nana* L. in Upper Teesdale. *New Phytologist*, *65*(3), 351–357.

Hynynen, J., Niemistö, P., Viherä-Aarnio, A., Brunner, A., Hein, S., & Velling, P. (2010). Silviculture of birch (*Betula pendula* Roth and *Betula pubescens* Ehrh.) in Northern Europe. *Forestry*, *83*(1), 103–119.

Ibrahim, K. M., Nichols, R. A, & Hewitt, G. M. (1996). Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity*, *77*, 282–291.

Ilut, D. C., Nydam, M. L., & Hare, M. P. (2014). Defining Loci in Restriction-Based Reduced Representation Genomic Data from Nonmodel Species : Sources of Bias and Diagnostics for Optimal Clustering, *2014*.

Ingvarsson, P. K., Hvidsten, T. R., & Street, N. R. (2016). Towards integration of population and comparative genomics in forest trees. *New Phytologist*, *212*(2), 338–344.

IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp. Available online: https://www.ipcc.ch/report/ar5/syr/

Ismail, S. A., Ghazoul, J., Ravikanth, G., Uma Shaanker, R., Kushalappa, C. G., & Kettle, C. J. (2012). Does long-distance pollen dispersal preclude inbreeding in tropical trees? Fragmentation genetics of *Dysoxylum malabaricum* in an agro-forest landscape. *Molecular Ecology*, *21*, 5484–5496.

Jeffrey, H. (1961). The Theory of Probability. *Oxford University Press, Oxford*. ISBN: 97 8019 85036 82

Jeffries, D. L., Copp, G. H., Handley, L. L., Olsén, K. H., Sayer, C. D., & Hänfling, B. (2015). Comparing RADseq and microsatellites to infer complex phylogeographic patterns , a real data informed perspective in the Crucian carp , *Carassius*. *Molecular Ecology*. 25(13). 2997-3018.

Jensen, J. D., & Foll, M. (2016). The past , present and future of genomic scans for selection. *Molecular Ecology*. 25(1). 1-4.

Jetz, W., Thomas, G. H., Joy, J. B., Redding, D. W., Hartmann, K., & Mooers, A. O. (2014). Global Distribution and Conservation of Evolutionary Distinctness in Birds. *Current Biology*, *24*(9), 919–930.

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405.

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., … Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*, 55–61.

Jones, T. A. (2013). When local isn't best. *Evolutionary Applications*, *6*(7), 1109–1118.

Joost, S., Bonin, A., Bruford, M. W., Despres, L., Conord, C., Erhardt, G., & Taberlet, P. (2007). A spatial analysis method (SAM) to detect candidate loci for selection: Towards a landscape genomics approach to adaptation. *Molecular Ecology*, *16*(18), 3955–3969.

Joost, S., Vuilleumier, S., Jensen, J. D., Schoville, S., Leempoel, K., Stucki, S., … Manel, S. (2013). Uncovering the genetic basis of adaptive change: On the intersection of landscape genomics and theoretical population genetics. *Molecular Ecology*, *22*(14), 3659–3665.

Jump, A. S., & Peñuelas, J. (2006). Genetic effects of chronic habitat fragmentation in a wind-pollinated tree. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(21), 8096–8100.

Kallio, P., Niemi, S., Sulkinoja, M., & & Valanne, T. (1983). The Fennoscandian birch and its evolution in the marginal forest zone. *Nordicana*, 47(10), 1-10.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281.

Kaplan, J. O., Krumhardt, K. M., & Zimmermann, N. (2009). The prehistoric and preindustrial deforestation of Europe. *Quaternary Science Reviews*, *28*(27-28), 3016–3034.

Karlsdóttir, L., Hallsdóttir, M., Thórsson, Æ. T., & Anamthawat-jónsson, K. (2009). Evidence of hybridisation between *Betula pubescens* and *B. nana* in Iceland during the early Holocene. *Review of Palaeobotany and Palynology*, *156*(3-4), 350–357.

Karlsson, H., Hornberg, G., Hannon, G., & Nordstrom, E. (2007). Long-term vegetation changes in the northern Scandinavian forest limit: a human impact climate synergy? *The Holocene*, *17*(2007), 37–49.

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*(12), 1225–1241.

Keller, S. R., Levsen, N., Olson, M. S., & Tiffin, P. (2012). Local adaptation in the flowering-time gene network of balsam poplar, *Populus balsamifera* L. *Molecular Biology and Evolution*, *29*(10), 3143–3152.

Kettle,C.J., Koh, L. P. (2014). Global Forest Fragmentation. *CAB International,* Wallingford. Available from www.cabi.org. ISBN: 9781780642031.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across *K*. *Molecular Ecology Resources*, *15*(5), 1179–1191.

Korner, C. (1998). A re-assessment of high elevation treeline psoitions and their explanation.

Oecologia, *115*(4), 445–459.

Körner, C. (2012). *Alpine Treelines: Functional Ecology of the Global High Elevation Tree Limits*. Springer Science & Business Media. ISBN: 978-3-0348-0395-3.

Koskela, J., Lefèvre, F., Schueler, S., Kraigher, H., Olrik, D. C., Hubert, J., … Ditlevsen, B. (2013). Translating conservation genetics into management: Pan-European minimum requirements for dynamic conservation units of forest tree genetic diversity. *Biological Conservation*. Elsevier Ltd.

Kramer, A. T., Ison, J. L., Ashley, M. V., & Howe, H. F. (2008). The Paradox of Forest Fragmentation Genetics. *Conservation Biology*, *22*(4), 878–885.

Kremer, A., Ronce, O., Robledo-Arnuncio, J. J., Guillaume, F., Bohrer, G., Nathan, R., … Schueler, S. (2012). Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters*, *15*(4), 378–392.

Kruzelnicki, J., & Fabiszewski, J. (2001). *Polska czerwona ksiega roslin [Polish Red Data Book of Plants]*. (K. Kazmierczakowa, R. & Zarzqcki, Ed.) (W. Szafer).

Kuhn, E., Lenoir, J., Piedallu, C., & Gégout, J.C. (2016). Early signs of range disjunction of sub-mountainous plant species: an unexplored consequence of future and contemporary climate changes. *Global Change Biology*, *22(6). 2094-2105.*

Kulju, K. K. M., Pekkinen, M., & Varvio, S. (2004). Twenty-three microsatellite primer pairs for *Betula pendula* (Betulaceae). *Molecular Ecology Notes*, *4*(3), 471–473.

Lacy, R. C. (1987). Loss of Genetic Diversity from Managed Populations: Interacting Effects of Drift, Mutation, Immigration, Selection, and Population Subdivision. *Conservation Biology*, *1*(2), 143–158.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Meth*, *9*(4), 357–359.

Lascoux, M., Palmé, A. E., Cheddadi, R., & Latta, R. G. (2004). Impact of Ice Ages on the genetic structure of trees and shrubs. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *359*(1442), 197–207.

Leimu, R., & Fischer, M. (2008). A meta-analysis of local adaptation in plants. *PLoS ONE*, *3*(12), 1–8.

Lenoir, J., Gégout, J. C., Marquet, P. A., de Ruffray, P., & Brisse, H. (2008). A significant upward shift in plant species optimum elevation during the 20th century. *Science (New York, N.Y.)*, *320*(5884), 1768–71.

Leonardi, S., Piovani, P., Scalfi, M., Piotti, A., Giannini, R., & Menozzi, P. (2012). Effect of habitat fragmentation on the genetic diversity and structure of peripheral populations of beech in central Italy. *Journal of Heredity*, *103*(3), 408–417.

Levin, D. A., Francisco-Ortega, J., & Jansen, R. K. (1996). Hybridization and the Extinction of Rare Plant Species. *Conservation Biology*, *10*(1), 10–16.

Li, J., Shoup, S., & Chen, Z. (2007). Phylogenetic relationships of diploid species of *Betula* (Betulaceae) inferred from DNA sequences of nuclear nitrate reductase. *Systematic Botany*, *32*, 357–365.

Li, Y.C., Korol, A. B., Fahima, T., Beiles, a, & Nevo, E. (2002). Microsatellites: genomic

distribution, putativa functions, and mutational mechanism: a review. *Molecular Ecology*, *11*, 253–256.

Lindner, M., Garcia-Gonzalo, J., Kolström, M., Green, T., Reguera, R., Maroschek, M., … Corona, P. (2008). Impacts of Climate Change on European Forests and Options for Adaptation. *Development*, 173.

Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298–299.

Lischer, L. E. and H. E. L. (2010). Arlequin suite V3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecolology*, *10*, 564–567.

Liu, C., Newell, G., & White, M. (2016). On the selection of thresholds for predicting species occurrence with presence-only data. *Ecology and Evolution*, *6*(1), 337–348.

Loarie, S. R., Duffy, P. B., Hamilton, H., Asner, G. P., Field, C. B., & Ackerly, D. D. (2009). The velocity of climate change. *Nature*, *462*, 1052–1055.

Lobo, J. M., Jiménez-valverde, A., & Real, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, *17*(2), 145–151.

Lowe, A. J., Cavers, S., Boshier, D., Breed, M. F., & Hollingsworth, P. M. (2015). The resilience of forest fragmentation genetics—no longer a paradox—we were just looking in the wrong place. *Heredity*, *115*(2), 97–99.

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2016). Breaking RAD: An evaluation of the utility of restriction site associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*. (early view)

Magri, D., Vendramin, G. G., Comps, B., Dupanloup, I., Geburek, T., Gömöry, D., … De Beaulieu, J. L. (2006). A new scenario for the Quaternary history of European beech populations: Palaeobotanical evidence and genetic consequences. *New Phytologist*, *171*(1), 199–221.

Maguire, K. C., Nieto-Lugilde, D., Fitzpatrick, M. C., Williams, J. W., & Blois, J. L. (2015). Modeling Species and Community Responses to Past, Present, and Future Episodes of Climatic and Ecological Change. *Annual Review of Ecology, Evolution, and Systematics*, *46*, 343–368.

Makela, E. (1998). The Holocene history of *Betula* at Lake Iilompolo, Inari Lapland, northeastern Finland. *Holocene*, *8*(1), 55–67.

Maliouchenko, O., Palmé, A. E., Buonamici, A., Vendramin, G. G., & 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*(9), 1601–1610.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, *20*(5), 229–237.

Mandák, B., Havrdová, A., Krak, K., Hadincová, V., Vít, P., Zákravský, P., & Douda, J. (2016). Recent similarity in distribution ranges does not mean a similar postglacial history: A phylogeographical study of the boreal tree species *Alnus incana* based on microsatellite and chloroplast DNA variation. *New Phytologist*, *210*(4), 1395–1407.

Manthey, J. D., Campbell, L. P., Saupe, E. E., Soberón, J., Hensz, C. M., Myers, C. E., … Barve, V. (2015). A test of niche centrality as a determinant of population trends and conservation status in threatened and endangered North American birds. *Endangered Species Research*, *26*(3), 201–208.

Manthey, J. D., & Moyle, R. G. (2015). Isolation by environment in White-breasted Nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: a landscape genomics approach. *Molecular Ecology*, 24(14). 3628-3638.

Mardon, D. K. (2003). Conserving montane willow scrub on Ben Lawers NNR. *Botanical Journal of Scotland*, *55*(1). 189-203.

Martins, W. S., Lucas, D. C. S., Neves, K. F. D. S., & Bertioli, D. J. (2009). WebSat - a web software for microsatellite marker development. *Bioinformation*, *3*(6), 282–283.

Mattioni, C., Angela Martin, M., Pollegioni, P., Cherubini, M., & Villani, F. (2013). Microsatellite markers reveal a strong geographical structure in European populations of *Castanea sativa* (Fagaceae): Evidence for multiple glacial refugia. *American Journal of Botany*, *100*(5), 951–961.

McGlone, M. S. (1996). When history matters: Scale, time, climate and tree diversity. *Global Ecology and Biogeography*, *5*(6), 309–314.

McKenney, D., Pedlar, J., Lawrance, K., Campbell, K., & Hutchinson, M. (2007). Potential impacts of climate change on the distribution of North American trees. *BioScience*, *57*(11), 939–948.

Meier, E. S., Lischke, H., Schmatz, D. R., & Zimmermann, N. E. (2012). Climate, competition and connectivity affect future migration and ranges of European trees. *Global Ecology and Biogeography*, *21*(2), 164–178.

Meirmans, P. G. (2015). Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology*, *24*(13), 3223–3231.

Meirmans, P. G., Goudet, J., & Gaggiotti, O. E. (2011). Ecology and life history affect different aspects of the population structure of 27 high-alpine plants. *Molecular Ecology*, *20*(15), 3144–3155.

Millar, C. I., Westfall, R. D., Delany, D. L., King, J. C., Graumlich, J., Millar, C. I., … King, J. C. (2004). Response of Subalpine Conifers in the Sierra Nevada, California, U.S.A., to 20th-Century Warming and Decadal Climate Variability. *Arctic, Antarctic, and Alpine Research*, *36*(2), 180–200.

Morin, P. A., Luikart, G., & Wayne, R. K. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution*, *19*(4), 208–216.

Park, T., Ganguly, S., Tommervik, H., Euskirchen, E.S., Hogda, K.A., Karlsen, S.R., Brovkin, V., Nemani, R. & Myneni, R.B. (2016). Changes in growing season duration and productivity of northern vegetation inferred from long-term remote sensing data. *Environmental Research Letters*, *11*(8), 84001.

Nakamura, M., Makoto, K., Tanaka, M., Inoue, T., Son, Y., & Hiura, T. (2016). Leaf flushing and shedding, bud and flower production, and stem elongation in tall birch trees subjected to increases in aboveground temperature. *Trees*, *30*(5), 1535–1541.

Narum, S. R., & Hess, J. E. (2011). Comparison of FST outlier tests for SNP loci under selection.

Molecular Ecology Resources, *11*(SUPPL. 1), 184–194.

Naydenov, K., Senneville, S., Beaulieu, J., Tremblay, F., & Bousquet, J. (2007). Glacial vicariance in Eurasia: mitochondrial DNA evidence from Scots pine for a complex heritage involving genetically distinct refugia at mid-northern latitudes and in Asia Minor. *BMC Evolutionary Biology*, *7*, 233.

Newton, A. C., Allnutt, T. R., Lowe, A. J., & Ennos, R. A. (1999). Molecular phylogeography, intraspecific variation and the conservation of tree species. *Trends in Ecology & Evolution*, *14*(4), 140–145.

Nichols, R. A., & Freeman, K. L. M. (2004). Using molecular markers with high mutation rates to obtain estimates of relative population size and to distinguish the effects of gene flow and mutation: A demonstration using data from endemic Mauritian skinks. *Molecular Ecology*, *13*(4), 775–787.

Nicotra, A. B., Atkin, O. K., Bonser, S. P., Davidson, A. M., Finnegan, E. J., Mathesius, U., … van Kleunen, M. (2010). Plant phenotypic plasticity in a changing climate. *Trends in Plant Science*, *15*(12), 684–692.

O'Connell, L. M., Mosseler, A., & Rajora, O. P. (2007). Extensive long-distance pollen dispersal in a fragmented landscape maintains genetic diversity in white spruce. *Journal of Heredity*, *98*(7), 640–645.

Olofsson, J., Oksanen, L., Callaghan, T., Hulme, P. E., Oksanen, T., & Suominen, O. (2009). Herbivores inhibit climate-driven shrub expansion on the tundra. *Global Change Biology*, *15*(11), 2681–2693.

Olson, M. S., Levsen, N., Soolanayakanahally, R. Y., Guy, R. D., Schroeder, W. R., Keller, S. R., & Tiffin, P. (2013). The adaptive potential of *Populus balsamifera* L. to phenology requirements in a warmer global climate. *Molecular Ecology*, *22*(5), 1214–1230.

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

Ossowski, S., Schneeberger, K., Lucas-Lledó, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., … Lynch, M. (2009). The Rate and Molecular Spectrum of Spontaneous Mutations in *Arabidopsis thaliana*. *Science*, *327*(5961), 92–94.

Pachauri, R. K., & Meyer, L. A. (2014). *IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Available from https://www.ipcc.ch/report/ar5/syr/

Palme, A. E., 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*(1), 167–178.

Palmé, A. E., Su, Q., Rautenberg, A., Manni, F., & Lascoux, M. (2003). Postglacial recolonization and cpDNA variation of silver birch, *Betula pendula*. *Molecular Ecology*, *12*(1), 201–212.

Paques, L. E. (2013). *Forest Tree Breeding in Europe: Current State-of-the-Art and Perspectives*. Springer Netherlands. (Eds: Pâques**,** L. E.) ISBN 978-94-007-6146-9.

Pardo-Diaz, C., Salazar, C., & Jiggins, C. D. (2015). Towards the identification of the loci of

adaptive evolution. *Methods in Ecology and Evolution*, *6*(4), 445–464.

Parmesan, C. (2006). Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics*, *37*(1), 637–669.

Parmesan, C., & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, *421*, 37–42.

Parolo, G., Rossi, G., & Ferrarini, A. (2008). Assessing ecological responses to environmental change using statistical models. *Journal of Applied Ecology*, *45*, 1410–1418.

Payseur, B. A., & Cutter, A. D. (2006). Integrating patterns of polymorphism at SNPs and STRs. *Trends in Genetics*, *22*(8), 424–429.

Pearson, R. G., Raxworthy, C. J., Nakamura, M., & Townsend Peterson, a. (2007). Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar. *Journal of Biogeography*, *34*(1), 102–117.

Persson, H., Widén, B., Andersson, S., & Svensson, L. (2004). Allozyme diversity and genetic structure of marginal and central populations of *Corylus avellan*a L. (Betulaceae) in Europe. *Plant Systematics and Evolution*, *244*(3-4), 157–179.

Petit, R. J. (2004). Biological invasions at the gene level. *Diversity and Distributions*, *10*(3), 159–165.

Petit, R. J., Aguinagalde, I., de Beaulieu, J.-L., Bittkau, C., Brewer, S., Cheddadi, R., … Vendramin, G. G. (2003). Glacial Refugia: Hotspots But Not Melting Pots of Genetic Diversity. *Science*, *300*(5625), 1563–1565.

Petit, R. J., Bialozyt, R., Garnier-Géré, P., & Hampe, A. (2004). Ecology and genetics of tree invasions: From recent introductions to Quaternary migrations. *Forest Ecology and Management*, *197*(1-3), 117–137.

Petit, R. J., & Hampe, A. (2006). Some Evolutionary Consequences of Being a Tree. *Annual Review of Ecology, Evolution, and Systematics*, *37*(1), 187–214.

Pfeifer-Meister, L., Bridgham, S. D., Little, C. J., Reynolds, L. L., Goklany, M. E., & Johnson, B. R. (2013). Pushing the limit: Experimental evidence of climate effects on plant range distributions. *Ecology*, *94*(10), 2131–2137.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3-4), 231–259.

Pihur, V., Datta, S., & Datta, S. (2009). RankAggreg, an R package for weighted rank aggregation. *BMC Bioinformatics*, *10*(1), 62.

Piotti, a, Leonardi, S., Buiteveld, J., Geburek, T., Gerber, S., Kramer, K., … Vendramin, G. G. (2012). Comparison of pollen gene flow among four European beech (*Fagus sylvatica* L.) populations characterized by different management regimes. *Heredity*, *108*(3), 322–331.

Piotti, A. (2009). The genetic consequences of habitat fragmentation: the case of forests. *iForest*, *2*(3), 75–76.

Polechová, J., & Barton, N. H. (2015). Limits to adaptation along environmental gradients. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(20), 6401–6406.

Potter, K. M., & Hargrove, W. W. (2012). Determining suitable locations for seed transfer under climate change: A global quantitative method. *New Forests*, *43*(5-6), 581–599.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

Priyam, A., Woodcroft, B. J., Rai, V., Munagala, A., Moghul, I., Ter, F., … Wurm, Y. (2015). Sequenceserver: a modern graphical user interface for custom BLAST databases. *bioRxiv*, 033142.

Prober, S. M., Byrne, M., McLean, E. H., Steane, D. A., Potts, B. M., Vaillancourt, R. E., & Stock, W. D. (2015). Climate-adjusted provenancing: a strategy for climate-resilient ecological restoration. *Frontiers in Ecology and Evolution*, *3*.

Provan, J., Beatty, G. E., Hunter, A. M., McDonald, R. A., McLaughlin, E., Preston, S. J., & Wilson, S. (2008). Restricted gene flow in fragmented populations of a wind-pollinated tree. *Conservation Genetics*, *9*(6), 1521–1532.

Provan, J., & Bennett, K. (2008). Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution*, *23*(10), 564–571.

Provan, J., & Maggs, C. A. (2012). Unique genetic variation at a species' rear edge is under threat from global climate change. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1726), 39–47.

R Development Core Team. (2014). R: A language and environment for statistical computing. *Foundation for Statistical Computing, Vienna, Austria.*

Randin, C.F., Engler, R., Normand, S., Zappa, M., Zimmermann, N.E., Pearman, P. B., … Guisan. (2009). Climate change and plant distribution: local models predict high-elevation persistence. *Global Change Biology*, *15*, 1557–1569.

Raymo, M. E. (1994). The Initiation of Northern Hemisphere Glaciation. *Annual Review of Earth Planet Sciences*, *22*, 353–383.

Reid, C. (1899). *The Origin of the British flora*. London: Dulau.

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

Rellstab, C., Zoller, S., Walthert, L., Lesur, I., Pluess, A. R., Graf, R., … Gugerli, F. (2016). Signatures of local adaptation in candidate genes of oaks (*Quercus* spp.) in respect to present and future climatic conditions. *Molecular Ecology*, 25(23). 5907–5924.

Renaut, S., Owens, G. L., & Rieseberg, L. H. (2014). Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. *Molecular Ecology*, *23*(2), 311–324.

Rendell, S., & Ennos, R. A. (2002). Chloroplast DNA diversity in *Calluna vulgaris* (heather) populations in Europe. *Molecular Ecology*, *11*(1), 69–78.

Rhymer, J. M., & Simberloff, D. (1996). Extinction by Hybridization and Introgression. *Annual Review of Ecology and Systematics*, *27*, 83–109.

Richards, M. (2008). Montane Scrub Survey 2008. Available from: http://www.highlandbirchwoods.co.uk/userfiles/file/publications/Scrubbers-

Bulletin/Bulletin8small.pdf

Rieseberg, L. H., Archer, M. a, & Wayne, R. K. (1999). Transgressive segregation, adaptation and speciation. *Heredity*, *83(4)*, 363–372.

Rius, M., & Darling, J. A. (2014). How important is intraspecific genetic admixture to the success of colonising populations? *Trends in Ecology & Evolution*, *29*(4), 233–242.

Robledo-Arnuncio, J. J., Collada, C., Al??a, R., & Gil, L. (2005). Genetic structure of montane isolates of *Pinus sylvestris* L. in a Mediterranean refugial area. *Journal of Biogeography*, *32*(4), 595–605.

Rolland, J., Lavergne, S., & Manel, S. (2015). Combining niche modelling and landscape genetics to study local adaptation: A novel approach illustrated using alpine plants. *Perspectives in Plant Ecology, Evolution and Systematics*, *17*(6), 491–499.

Rounsevell, M. D. A., Reginster, I., Araújo, M. B., Carter, T. R., Dendoncker, N., Ewert, F., … Tuck, G. (2006). A coherent set of future land use change scenarios for Europe. *Agriculture, Ecosystems and Environment*, *114*(1), 57–68.

Rousset, F. (2008). GENEPOP’007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, *8*(1), 103–106.

Royle, J. A., Chandler, R. B., Yackulic, C., & Nichols, J. D. (2012). Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. *Methods in Ecology and Evolution*, *3*(3), 545–554.

Rusanen, M., Vakkari, P., & Blom, A. (2003). Genetic structure of *Acer platanoides* and *Betula pendula* in northern Europe. *Canadian Journal of Forest Research*, *33*(6), 1110–1115.

Ryan, D., Farrell, E. P., Fu, E., Andersson, F., & Hu, R. (2000). European forest ecosystems: building the future on the legacy of the past, *132(1)*. 5-20.

Salmela, M. J., Cavers, S., Wachowiak, W., Cottrell, J. E., Iason, G. R., & Ennos, R. A. (2010). Understanding the evolution of native pinewoods in Scotland will benefit their future management and conservation. *Forestry*, *83*(5), 535–545.

Santini, A., Ghelardini, L., De Pace, C., Desprez-Loustau, M. L., Capretti, P., Chandelier, A., … Stenlid, J. (2013). Biogeographical patterns and determinants of invasion by forest pathogens in Europe. *New Phytologist*, *197*(1), 238–250.

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews. Genetics*, *14*(11), 807–20.

Savolainen, O., Pyhäjärvi, T., & Knürr, T. (2007). Gene Flow and Local Adaptation in Trees. *Annual Review of Ecology, Evolution, and Systematics*, *38*(1), 595–619.

Scherrer, D., & Körner, C. (2011). Topographically controlled thermal-habitat differentiation buffers alpine plant diversity against climate warming. *Journal of Biogeography*, *38*(2), 406–416.

Schönswetter, P., Paun, O., Tribsch, A., & Niklfeld, H. (2003). Out of the Alps: Colonization of Northern Europe by East Alpine populations of the Glacier Buttercup *Ranunculus glacialis* L. (Ranunculaceae). *Molecular Ecology*, *12*(12), 3373–3381.

Schönswetter, P., Stehlik, I., Holderegger, R., & Tribsch, A. (2005). Molecular evidence for glacial refugia of mountain plants in the European Alps. *Molecular Ecology*, *14*(11), 3547–

3555.

Schweitzer, J. A., Martinsen, G. D., & Whitham, T. G. (2002). Cottonwood hybrids gain fitness traits of both parents: A mechanism for their long-term persistence? *American Journal of Botany*, *89*(6), 981–990.

Scott, M. (2000). *Natural Heritage management series: Montane Scrub*. Scottish Natural Heritage. Available from: http://www.snh.org.uk/pdfs/publications/heritagemanagement/montanescrub.pdf

Searcy, C. A., & Shaffer, H. B. (2016). Do Ecological Niche Models Accurately Identify Climatic Determinants of Species Ranges? *The American Naturalist*, *187:4*, 423–435.

Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution*, *68*(1), 1–15.

Shaffer, H. B., Fisher, R. N., & Davidson, C. (1998). The role of natural history collections in documenting species declines. *Trends in Ecology & Evolution*, *13*(1), 27–30.

Siol, M., Wright, S. I., & Barrett, S. C. H. (2010). The population genomics of plant adaptation. *New Phytologist*, *188*(2), 313–332.

Skrede, I., Borgen, L., & Brochmann, C. (2009). Genetic structuring in three closely related circumpolar plant species: AFLP versus microsatellite markers and high-arctic versus arctic–alpine distributions. *Heredity*, *102120*, 293–302.

Smart, J. C. R., Ward, A. I., & White, P. C. L. (2004). Monitoring woodland deer populations in the UK: an imprecise science. *Mammal Review*, *34*(1-2), 99–114.

Smith, S. J., Edmonds, J., Hartin, C. A., Mundra, A., & Calvin, K. (2015). Near-term acceleration in the rate of temperature change. *Nature Climate Change*, *5*(4), 333–336.

Smout, T., MacDonald, A.R., & Watson, F. (2005). *A History of the Native Woodlands of Scotland, 1500-1920*. Edinburgh University Press, 2005, 434p.

Soyrinki, N. (1939). Studien uber die generative und vegetative Vermehrung der Samenpflanzen in der alpinen Vegetation Petsamo-Lapplands. *Ann. Bot. Soc. Zool. Bot. Fenn.*, 14, 1–404.

Springate, D. A., & Kover, P. X. (2014). Plant responses to elevated temperatures: A field study on phenological sensitivity and fitness responses to simulated climate warming. *Global Change Biology*, *20*(2), 456–465.

St Clair, J. B., & Howe, G. T. (2007). Genetic maladaptation of coastal Douglas-fir seedlings to future climates. *Global Change Biology*, *13*(7), 1441–1454.

Stanton-Geddes, J., Tiffin, P., & Shaw, R. G. (2012). Role of climate and competitors in limiting fitness across range edges of an annual plant. *Ecology*, *93*(7), 1604–1613.

Stephenson, N. L. (1990). Climatic Control of Vegetation Distribution: The Role of the Water Balance. *American Naturalist*. 135(5). 649-670.

Stewart, J. R., & Lister, A. M. (2001). Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology and Evolution*, *16*(11), 608–613.

Stewart, A., Pearman, D.A., & Preston, C. D. (1994). *Scarce plants in Britain*. Peterborough: Joint Nature Conservation Committee. ISBN 1 87370 166 7.

Suarez-Gonzalez, A., Hefer, C. A., Christe, C., Corea, O., Lexer, C., Cronk, Q. C. B., & Douglas, C. J. (2016). Genomic and functional approaches reveal a case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. trichocarpa* (black cottonwood). *Molecular Ecology*, 25(11). 2427-2442.

Svenning, J.-C. (2003). Deterministic Plio-Pleistocene extinctions in the European cool-temperate tree flora. *Ecology Letters*, *6*(7), 646–653.

Szpiech, Z. A., Jakobsson, M., & Rosenberg, N. A. (2008). ADZE: A rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, *24*(21), 2498–2504.

Tanentzap, A. J., Zou, J., & Coomes, D. A. (2013). Getting the biggest birch for the bang: restoring and expanding upland birchwoods in the Scottish Highlands by managing red deer. *Ecology and Evolution. 3*(7). 1890–1901.

Tarkesh, M., & Gottfried, J. (2012). Comparison of six correlative models in predictive vegetation mapping on a local scale. *Environmental and Ecological Statistics*, *19*(3), 437–457.

Taylor, S. A., Larson, E. L., & Harrison, R. G. (2015). Hybrid zones: windows on climate change. *Trends in Ecology & Evolution*, *30*(7), 398–406.

Team, Rs. (2015). RStudio: Integrated Development for R. *RStudio, Inc., Boston, MA*.

Th.Thórsson, A. E., 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*(11), 2098–2110.

Thomas, C. D. (2015). Rapid acceleration of plant speciation during the Anthropocene. *Trends in Ecology and Evolution*, *30*(8), 448–455.

Thompson, D.B.A., & Brown, A. (1992). Biodiversity in montane Britain: habitat variation, vegetation diversity and some objectives for conservation. *Biodiversity and Conservation*, *1*(3), 179–208.

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

Thórsson, A. T., Pálsson, S., Sigurgeirsson, a, & Anamthawat-Jónsson, K. (2007). Morphological variation among *Betula nana* (diploid), *B. pubescens* (tetraploid) and their triploid hybrids in Iceland. *Annals of Botany*, *99*(6), 1183–93.

Thuiller, W., Albert, C. H., Dubuis, A., Randin, C., & Guisan, A. (2010). Variation in habitat suitability does not always relate to variation in species' plant functional traits. *Biology Letters*, *6*(1), 120–3.

Thuiller, W., Lavorel, S., Araujo, M. B., Sykes, M. T., & Prentice, I. C. (2005). Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences*, *102*(23), 8245–8250.

Thuiller, W., Lavorel, S., Sykes, M. T., & Araújo, M. B. (2006). Using niche-based modelling to assess the impact of climate change on tree functional diversity in Europe. *Diversity and Distributions*, *12*(1), 49–60.

Tingley, M. W., & Beissinger, S. R. (2009). Detecting range shifts from historical species occurrences: new perspectives on old data. *Trends in Ecology and Evolution*, *24*(11), 625–633.

Tipping, R. (1994). The form and fate of Scotland's woodlands. *Proceedings of the Society of Antiquaries of Scotland*, *124*, 1–54.

Truong, C., Palmé, a. E., & 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.

Truong, C., Palme, A. E., Felber, F., & Naciri-Graven, Y. (2005). Isolation and characterization of microsatellite markers in the tetraploid birch, *Betula pubescens ssp. tortuosa*. *Molecular Ecology Notes*, *5*(1), 96–98.

Tsuda, Y., Semerikov, V., Sebastiani, F., Vendramin, G. G., & Lascoux, M. (2016a). Multispecies genetic structure and hybridization in the *Betula* genus across Eurasia. *Molecular Ecology*. 26(2). 589-605.

Tuley, G. (1973). A taxonomic history of the British birch tree. *Transactions of the Botanical Society of Edinburgh*, *42*(1), 31–41.

Tuskan, G. a., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, *313*(5793), 1596.

Tzedakis, P. C., Emerson, B. C., & Hewitt, G. M. (2013). Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends in Ecology & Evolution*, *28*(12), 696–704.

Udupa, S. M., & Baum, M. (2001). High mutation rate and mutational bias at (TAA) in microsatellite loci in chickpea (*Cicer arietinum* L.). *MGG Molecular Genetics and Genomics*, *265*(6), 1097–1103.

Valjakka, M., Aronen, T., Kangasjarvi, J., Vapaavuori, E., & Haggman, H. (2000). Genetic transformation of silver birch (*Betula pendula*) by particle bombardment. *Tree Physiology*, *20*(9), 607–613.

Vallone, P. M., & Butler, J. M. (2004). AutoDimer: A screening tool for primer-dimer and hairpin structures. *BioTechniques*, *37*(2), 226–231.

Van Bers, N. E. M., Van Oers, K., Kerstens, H. H. D., Dibbits, B. W., Crooijmans, R. P. M. A., Visser, M. E., & Groenen, M. A. M. (2010). Genome-wide SNP detection in the great tit *Parus major* using high throughput sequencing. *Molecular Ecology*, *19*(SUPPL. 1), 89–99.

Via, S., & Lande, R. (1985). Genotype-Environment Interaction and the Evolution of Phenotypic Plasticity. *Evolution*, *39*(3), 505–522.

Vila, M., Weber, E., & Antonio, C. M. D. (2000). Conservation implications of invasion by plant hybridization. *Biological Invasions*, *2*(3), 207–217.

Vinceti, B., Loo, J., Gaisberger, H., van Zonneveld, M. J., Schueler, S., Konrad, H., … Geburek, T. (2013). Conservation Priorities for Prunus africana Defined with the Aid of Spatial Analysis of Genetic Data and Climatic Variables. *PLoS ONE*, *8*(3).

Volkmann, L., Martyn, I., Moulton, V., Spillner, A., & Mooers, A. O. (2014). Prioritizing populations for conservation using phylogenetic networks. *PLoS ONE*, *9*(2).

Vranckx, G., Jacquemyn, H., Muys, B., & Honnay, O. (2012). Meta-Analysis of Susceptibility of Woody Plants to Loss of Genetic Diversity through Habitat Fragmentation. *Conservation Biology*, *26*(2), 228–237.

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

Walck, J. L., Hidayati, S. N., Dixon, K. W., Thompson, K., & Poschlod, P. (2011). Climate change and plant regeneration from seed. *Global Change Biology*, *17*(6), 2145–2161.

Wang, N., Borrell, J. S., Bodles, W. J. a, Kuttapitiya, A., Nichols, R. a., & Buggs, R. J. a. (2014). Molecular footprints of the Holocene retreat of dwarf birch in Britain. *Molecular Ecology*, *23*(11), 2771–2782.

Wang, N., Mcallister, H. A., Bartlett, P. R., & Buggs, R. J. A. (2016). Molecular phylogeny and genome size evolution of the genus *Betula* (Betulaceae). *Annals of Botany*, *117*(6), 1023–1035.

Wang, N., Thomson, M., Bodles, W. J. a, Crawford, R. M. M., Hunt, H. V., Featherstone, A. W., … Buggs, R. J. a. (2013). Genome sequence of dwarf birch (*Betula nana*) and cross-species RAD markers. *Molecular Ecology*, *22*(11), 3098–3111.

Wang, T., Wang, Z., Xia, F., & Su, Y. (2016). Local adaptation to temperature and precipitation in naturally fragmented populations of *Cephalotaxus oliveri*, an endangered conifer endemic to China. *Scientific Reports*, *6*, 25031.

Ward, M., Dick, C. W., Gribel, R., & Lowe, A. J. (2005). To self, or not to self... a review of outcrossing and pollen-mediated gene flow in neotropical trees. *Heredity*, *95*(4), 246–254.

Warren, D. L., Glor, R. E., & Turelli, M. (2008). Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution*, *62*(11), 2868–2883.

Weber, M. M., Stevens, R. D., Diniz-Filho, J. A. F., & Grelle, C. E. V. (2016). Is there correlation between abundance and environmental suitability derived from ecological niche modeling ? A meta-analysis. *Ecography*, *39*, 1–12.

Weis, I. M., & Hermanutz, L. A. (1993). Pollination dynamics of Arctic dwarf birch (*Betula glandulosa*: Betulaceae) and its role in the loss of seed production. *American Journal of Botany*, *80*(9), 1021–1027.

Wesołowski, T., & Rowiński, P. (2006). Tree defoliation by winter moth *Operophtera brumata* L. during an outbreak affected by structure of forest landscape. *Forest Ecology and Management*, *221*(1-3), 299–305.

Whitlock, M. C. (1992). Temporal Fluctuations in Demographic Parameters and the Genetic Variance among Populations. *Evolution*, *46*(3), 608–615.

Whittet, R., Cavers, S., Cottrell, J., & Ennos, R. (2016). Seed sourcing for woodland creation in an era of uncertainty: an analysis of the options for Great Britain. *Forestry*, 1–11.

Whittet, R., Cottrell, J., Cavers, S., Pecurul, M., & Ennos, R. (2016). Supplying trees in an era of environmental uncertainty: Identifying challenges faced by the forest nursery sector in Great Britain. *Land Use Policy*, *58*, 415–426.

Wickham, H. (2009). ggplot2. Elegant graphics for data analysis. *Springer*. ISBN 978-0-387-

98141-3

Wilczek, A. M., Cooper, M. D., Korves, T. M., & Schmitt, J. (2014). Lagging adaptation to warming climate in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(22), 7906–13.

Williams, M., & Dumroese, K. (2013). Preparing for Climate Change: Forestry and Assisted Migration. *Journal of Forestry*, *111*, 287–297.

Willis, K. J., Rudner, E., & Sümegi, P. (2000). The Full-Glacial Forests of Central and Southeastern Europe. *Quaternary Research*, *53*(2), 203–213.

Wittmann, M. E., Barnes, M. A., Jerde, C. L., Jones, L. A., & Lodge, D. M. (2016). Confronting species distribution model predictions with species functional traits. *Ecology and Evolution*, *6*(4), 873–879.

Wolf, D. E., Takebayashi, N., & Rieseberg, L. H. (2001). Predicting the risk of extinction through hybridization. *Conservation Biology*, *15*(4), 1039–1053.

Young, A., Boyle, T., & Brown, T. (1996). The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology & Evolution*, *11*(10), 413–418.

Zhang, Z., & Gerstein, M. (2003). Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes. *Nucleic Acids Research*, *31*(18), 5338–5348.

Zhu, K., Woodall, C. W., & Clark, J. S. (2012). Failure to migrate: Lack of tree range expansion in response to climate change. *Global Change Biology*, *18*(3), 1042–1052.

Zohren, J., Wang, N., Kardailsky, I., Borrell, J. S., Joecker, A., Nichols, R. A., & Buggs, R. J. A. (2016). Unidirectional diploid-tetraploid introgression among British birch trees with shifting ranges shown by RAD markers. *Molecular Ecology*. *25*. 2413-2426.