

# Teasing apart the joint effect of demography and natural selection in the birth of a contact zone

Lili Li<sup>1\*</sup>, Pascal Milesi<sup>1\*</sup> , Mathieu Tiret<sup>1</sup> , Jun Chen<sup>1,2</sup> , Janek Sendrowski<sup>1</sup>, John Baison<sup>3</sup>,  
Zhi-qiang Chen<sup>3</sup> , Linghua Zhou<sup>3</sup>, Bo Karlsson<sup>4</sup>, Mats Berlin<sup>5</sup> , Johan Westin<sup>6</sup>, Maria Rosario Garcia-Gil<sup>3</sup> ,  
Harry X. Wu<sup>3,7</sup>  and Martin Lascoux<sup>1</sup> 

<sup>1</sup>Program in Plant Ecology and Evolution, Department of Ecology and Genetics, EBC and SciLife Lab, Uppsala University, 75236 Uppsala, Sweden; <sup>2</sup>College of Life Sciences, Zhejiang University, Hangzhou Zhejiang 310058, China; <sup>3</sup>Department Forest Genetics and Plant Physiology, Umeå Plant Science Centre, Swedish University of Agricultural Sciences, Umeå SE-90183, Sweden; <sup>4</sup>Skogforsk, Ekebo 2250, SE-268 90, Svalöv, Sweden; <sup>5</sup>Skogforsk, Uppsala Science Park, 751 83 Uppsala, Sweden; <sup>6</sup>Unit for Field-Based Forest Research, Swedish University of Agricultural Sciences, SE-922 91, Vindeln, Sweden; <sup>7</sup>CSIRO National Collection Research Australia, Black Mountain Laboratory, Canberra, ACT 2601, Australia

## Summary

Authors for correspondence:

Martin Lascoux

Email: martin.lascoux@ebc.uu.se

Pascal Milesi

Email: pascal.milesi@ebc.uu.se

Received: 18 May 2022

Accepted: 23 August 2022

*New Phytologist* (2022) **236**: 1976–1987

doi: 10.1111/nph.18480

**Key words:** climate change, contact zone, demography, Last Glacial Maximum, local adaptation, natural selection, *Picea abies*.

- Vast population movements induced by recurrent climatic cycles have shaped the genetic structure of plant species. During glacial periods species were confined to low-latitude refugia from which they recolonized higher latitudes as the climate improved. This multipronged recolonization led to many lineages that later met and formed large contact zones.
- We utilize genomic data from 5000 *Picea abies* trees to test for the presence of natural selection during recolonization and establishment of a contact zone in Scandinavia.
- Scandinavian *P. abies* is today made up of a southern genetic cluster originating from the Baltics, and a northern one originating from Northern Russia. The contact zone delineating them closely matches the limit between two major climatic regions. We show that natural selection contributed to its establishment and maintenance. First, an isolation-with-migration model with genome-wide linked selection fits the data better than a purely neutral one. Second, many loci show signatures of selection or are associated with environmental variables. These loci, regrouped in clusters on chromosomes, are often related to phenology.
- Altogether, our results illustrate how climatic cycles, recolonization and selection can establish strong local adaptation along contact zones and affect the genetic architecture of adaptive traits.

## Introduction

Tree species do not stand still and, although they are sessile organisms, their ranges expand and contract. During their evolution, most species ranges were forced into cycles of contraction and expansion by the alternation of contrasting climatic periods (Bennett, 1997). Some species disappeared while others survived. Among the survivors, demographic cycles were generally accompanied by separation into divergent genetic clusters, reflecting the location of the surviving populations during periods of harsh climate. As climatic conditions progressively improved, species regained lost ground and, often, the lineages created by the separation into different refugia met, thereby creating large contact zones. These contact zones, many of which are fairly young as they were established after the Last Glacial Maximum (LGM, c. 18 000 yr ago), can be viewed as large-scale competition experiments between divergent lineages and are therefore a rich source

of information on the interplay between demography and selection in shaping the genetic structure of species (Johannesson *et al.*, 2020).

In contrast to hybrid zones where the two species will generally experience some level of reproductive barrier (Moran *et al.*, 2021), contact zones are expected to be rapidly homogenized by gene flow, unless selection is present and acts swiftly. Stabilizing selection with different optima across an ecological transition (ecotone) was shown to be maintaining contact zones in a variety of species (Alberto *et al.*, 2013; Johannesson *et al.*, 2020). One can expect such contact zones to settle on ecological transitions even though the primary contact might have happened elsewhere. In turn, the position of the contact zones can inform us on ecological transitions and their movements with climate change (Wielstra, 2019, 2021). Stabilizing selection, along with gene flow, can lead to specific genetic architectures, where genes under selection tend to be clustered in a few specific regions of the genome, instead of being randomly distributed (Yeaman & Whitlock, 2011; Yeaman, 2013). Because of the relative youth of the contact zones and despite migration and drift,

\*These authors contributed equally to the present study and are listed in alphabetical order.

the genes under selection are expected to remain polymorphic (Savolainen *et al.*, 2013), and thus contribute to the high evolutionary potential of these populations. More generally, recent studies have shown that even long-lived species such as forest trees are able to respond rapidly to abrupt environmental changes (Dauphin *et al.*, 2021; Saleh *et al.*, 2022). The study of recent contact zones can therefore provide much needed information for forecasting species reactions to current global climate change.

The recolonization of Scandinavia after the LGM led to a postulated zone of postglacial contact in many species (Hewitt, 2000; Tollefsrud *et al.*, 2008). Such a contact zone is, for instance, observed in humans (Günther *et al.*, 2018; Peter *et al.*, 2020), aspens (De Carvalho *et al.*, 2010), brown bears (Bray *et al.*, 2013) and rodents (Jaarola *et al.*, 1999). In all these organisms, the contact zone has been initially interpreted as the meeting point between the two main lineages that recolonized Scandinavia after the LGM.

Norway spruce (*Picea abies* (L.) H. Karst.) is one of the most common boreal tree species. Since the seminal study of Lagercrantz & Ryman (1990), a large number of studies have outlined the most salient features of the demographic history of Norway spruce (Bucci & Vendramin, 2000; Vendramin *et al.*, 2000; Heuertz *et al.*, 2006; Tollefsrud *et al.*, 2008, 2009, 2015; Chen *et al.*, 2012, 2019; Tsuda *et al.*, 2016; Fagernäs, 2017; Wang *et al.*, 2020). Current populations emerged from three main glacial refugia located in the Alps, the Carpathians and the Russian plains. A recent study indicated that these main lineages did not evolve independently, but were instead influenced by contact zones (Chen *et al.*, 2019). The nature of these contact zones remains to be elucidated: they could simply be a reflection of past distribution shifts, correspond to ecological zones and be associated with local adaptation, or be the result of both. In Norway spruce, as is often the case for forest trees, local adaptation is strong, in particular along latitudinal gradients (Savolainen *et al.*, 2007, 2013). As expected under strong local adaptation, the pattern of differentiation of genotypic, phenotypic and environmental variables at the site of origin of the trees are highly correlated (Milesi *et al.*, 2019). Patterns of local adaptation can, for instance, be linked to changes in phenology. For example, growth cessation follows a latitudinal cline and is controlled by some major candidate genes (e.g. *FTL2*, Chen *et al.*, 2012, 2014). The initial studies were based on a handful of candidate genes, but more recent studies relying on a much larger number of markers (e.g. Milesi *et al.*, 2019; Chen *et al.*, 2021) indicate that phenology-related traits have a polygenic inheritance with loci involved in local adaptation distributed across the genome.

In the present study we sequenced all individuals from the base population of the Swedish *P. abies* breeding program using exome capture (4769 individuals, 12 Mb sequenced, > 500 000 single nucleotide polymorphisms (SNPs)), generating an unprecedentedly large and dense sampling along a latitudinal gradient ranging from *c.* 55°N to *c.* 67°N (Fig. 1a). Through analysis of these data, we characterized the population genetic structure at a global scale and identified its local drivers along the latitudinal post-LGM contact zone in Sweden. More specifically, we tested for isolation-by-distance, identified barriers to gene flow and tested whether they reflect physical or environmental

barriers or simply historical contingencies. We show that the northern range of Norway spruce is divided into two main genetic clusters that significantly match the two main climatic zones of the region. Coalescent simulations and Approximate Bayesian Computation allowed the rejection of a purely neutral divergence model between the two main clusters. Furthermore, genome scans indicated that adaptive loci were localized in clusters distributed across the 12 linkage groups of the Norway spruce genome. These clusters correspond to genomic areas of high genetic differentiation and are associated with environmental variables discriminating the two climatic zones. The current distribution of genetic diversity in Norway spruce across Scandinavia therefore appears to be the result of both ancient demographic processes associated with the main climatic cycles and rapid adaptation to local conditions after the LGM.

## Materials and Methods

### Sample collection

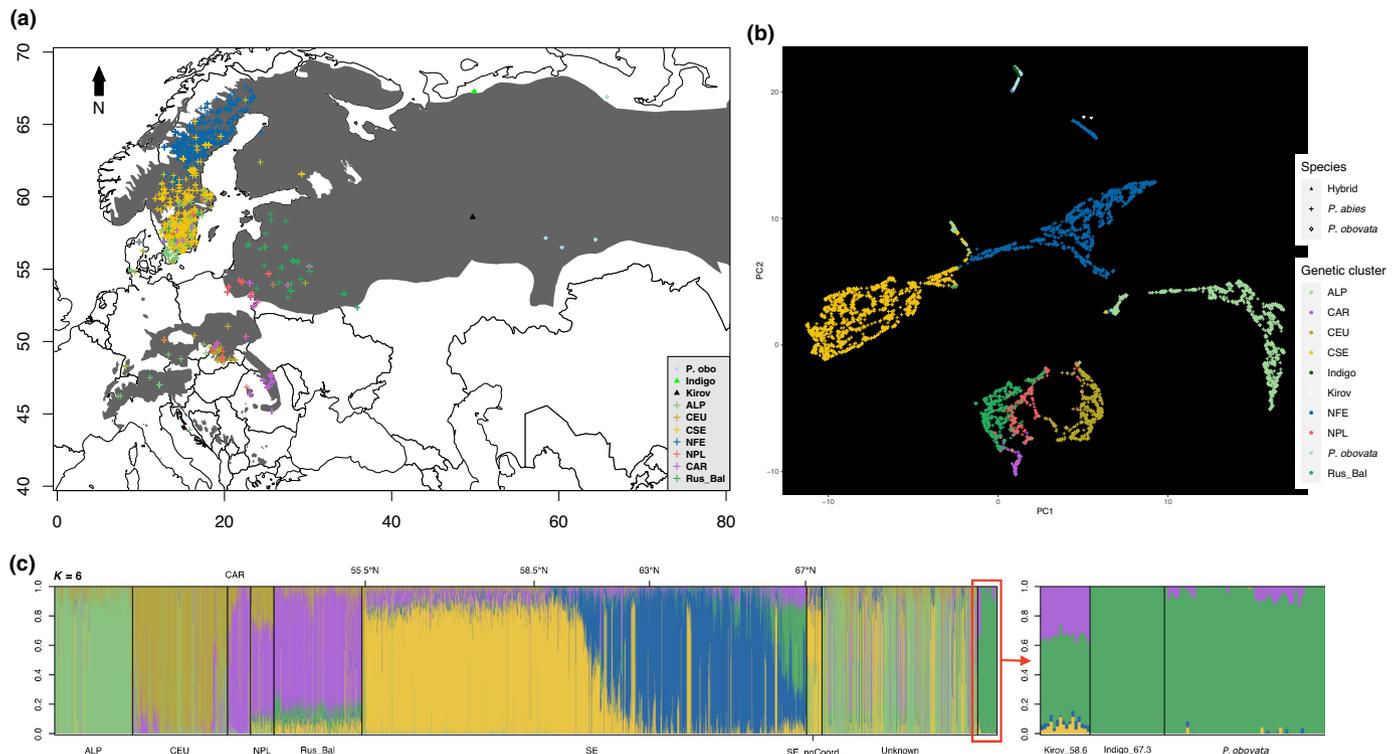
The study was based on 4769 trees (Fig. 1). Most individuals (4607) were 'plus trees' (trees of outstanding phenotype) collected in natural populations to establish the breeding population. They were sampled in Skogforsk (The Forestry Research Institute of Sweden) trials across Sweden. Geographic information on the origin of 873 of those trees was missing (Table S1). An additional 162 individuals were also collected in natural populations across the range of Norway spruce (40 individuals from the hybrid zone between *P. abies* and *P. obovata*, 69 unadmixed *P. abies* and 53 unadmixed *P. obovata*) (Chen *et al.*, 2019). These trees were genotyped (Chen *et al.*, 2019, BioProject PRJNA511374; and Chen *et al.*, 2021, BioProject PRJNA731384) using an exome capture strategy (*c.* 12 MB sequenced with 40 018 diploid 120 bp probes spread over 26 219 *P. abies* genes; Vidalis *et al.*, 2018). For further details see Methods S1.

### Single nucleotide polymorphism calling

Raw reads were mapped to the *P. abies* genome reference v.1.0 (Nystedt *et al.*, 2013) and SNPs were identified using HAPLOTYPECALLER v.3.6 (Li *et al.*, 2009) and quality filtered. Individuals with more than 50% missing data were also removed ( $N=282$ ). The filtered dataset included 4508 individuals and 504 110 SNPs. Those SNPs were annotated based on the most recent genome annotation available for *P. abies* (v.1.0, <http://congenie.org/>). For further details see Methods S1.

### Population structure and genotype assignment

For population structure analyses, sites in high linkage disequilibrium ( $r^2 > 0.2$ ) as well as singletons were removed using PLINK v.1.9 (Chang *et al.*, 2015). Among the remaining 302 793 SNPs, 155 211 putatively neutral SNPs (i.e. synonymous sites, or sites within introns and intergenic regions) were kept for demographic analyses. Population structure was first characterized using a principal component analysis (EIGENSOFT, v.7.2.0 with default parameters, <https://github.com>.



**Fig. 1** Population genetic structure of *Picea abies*. (a) Sampling location of *P. abies* (plus signs): light green, Alpine (ALP); brown, Central Europe (CEU); purple, Carpathian (CAR); red, northern Poland (NPL); dark green, Russia-Baltics (Rus\_Bal); yellow, southern Fennoscandia (CSE); and dark blue, northern Fennoscandia (NFE). *Picea obovata* (diamonds, light blue) and hybrids (Indigo and Kirov) (black triangles). The shaded area corresponds to the distribution range of *P. abies* and *P. obovata*. (b) UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction) bidimensional plots; colors are the same as for (a). (c) ADMIXTURE plot for  $K=6$ . Samples from the same geographic origin were grouped and Swedish samples (SE) were ordered by latitude. Colors represent different ancestry components. The red box designates the *P. obovata* samples and the part on the right of the arrow a magnified view of it.

com/DreichLab/EIG, Galinsky *et al.*, 2016) and UMAP (Diaz-Papkovich *et al.*, 2019). Trees with unknown geographic origin were assigned to a genetic cluster using Random Forest classification as in Chen *et al.* (2019). We also analyzed population structure with ADMIXTURE v.1.3 (Alexander *et al.*, 2009) and calculated pairwise fixation indices (Hudson's estimator of  $F_{ST}$ ; Hudson *et al.*, 1992) between *P. obovata*, admixed *P. abies*  $\times$  *P. obovata* populations and the *P. abies* genetic clusters defined through the UMAP analysis.

### Spatialized analyses of genetic variation

For the following analyses, only trees that were of confirmed Swedish origin (based on genetic clustering) and with known geographic coordinates were considered ( $N=1758$ ). To account for the continuous distribution of Norway spruce and to identify barriers to gene flow, we used CONSTRUCT v.1.03 (Bradburd *et al.*, 2018). To visualize the variation in effective migration rate across Sweden we fitted the data to a model of isolation-by-distance with EEMS v.0.0.9000 (Petkova *et al.*, 2015). Finally, we also quantified the pattern of isolation by distance (IBD) by regressing a function of pairwise  $F_{ST}$ , namely  $F_{ST}/(1-F_{ST})$ , over the logarithm of the distance between pairs of populations (Rousset, 1997). According to Rousset (1997), the inverse of the slope of the regression provides an indirect estimate of the neighborhood size ( $N_s$ ).

### Contribution of linked selection to the contact zone

To test whether natural selection contributed to the establishment and maintenance of the contact zone in Sweden between the northern (hereafter NFE) and the southern (hereafter CSE) genetic domains, we used the program DILS (Fraïsse *et al.*, 2021), which implements a coalescent-based approach to simulate a pre-defined set of isolation with migration models, and which calculates their posterior probabilities with an Approximate Bayesian Computation (ABC) approach (for detailed explanations see 'Contribution of linked selection to the contact zone' in Methods S1). Briefly, in the presence of linked selection, one would expect a larger variance in effective population size,  $N_e$ , among loci than under a strictly isolation with migration model. Considering that the genetic distance between individuals from NFE and CSE might influence demographic inferences (e.g. admixed individuals have a different history than unadmixed ones), we created three different datasets, each made up of 20 individuals (10 from each genetic domain). The individuals were selected as representative of the Northern and the Southern genetic cluster (ADMIXTURE,  $K=6$ ), but with different levels of admixture between the two clusters. The admixture index ( $A_{ind}$ ) is defined as  $A_{ind} = A_{NFE}/(A_{NFE} + A_{CSE})$  where  $A_x$  stands for the ancestry coefficient of cluster  $x$ . As the genetic distance varies with the geographic distance to the contact zone, we defined three datasets.

The ‘close’ dataset comprises individuals with  $0.37 < A_{\text{ind}} < 0.63$ , the ‘intermediate’ one, individuals with  $0.14 < A_{\text{ind}} < 0.36$  or  $0.64 < A_{\text{ind}} < 0.87$ , and the ‘far’ dataset individuals with  $A_{\text{ind}} < 0.04$  or  $A_{\text{ind}} > 0.96$ . Each of these datasets was used as input for DILS (Notes S2). We also used forward simulations to test whether a neutral model could explain the data and reached the same conclusion as with DILS (Notes S2).

### Testing for local adaptation

First, to assess whether the contact zone between the main genetic clusters corresponded to a shift in abiotic conditions across Sweden, we defined climatic zones based on 19 bioclimatic variables (Chelsa database v.1.2, <http://chelsa-climate.org>, 30 arc-second resolution). We then tested for the concordance between climatic and genetic variation along the contact zone. To do so we investigated the strength of association between the coordinates on the two first principal components of the principal components analysis (PCA) based on climatic variables at tree locations with corresponding admixture index ( $A_{\text{ind}}$ ) of the trees (see Notes S1 for more details). Second, we scanned our genomic data for loci showing extreme differentiation pattern using either an individual-based approach (PCADAPT v.4.3.2 R package, Luu *et al.*, 2017; Privé *et al.*, 2020), or allelic frequency variation between the 47 populations (BAYENV2,  $\chi^2$  score, Coop *et al.*, 2010; Günther & Coop, 2013). Third, we used BAYENV2 and LFMM2 (Caye *et al.*, 2019) to detect significant genotype–environment associations (GEAs). Genome scans and GEA analyses were conducted on a subset of 142 765 SNPs with an minimum allele frequency  $> 0.05$  and stringent criteria were used to control for false-positive detection risk associated with multiple testing (details are provided ‘Testing for local adaptation’ in Methods S1).

### Candidate genes putative functions and genetic mapping

The genes putatively involved in local adaptation were tested for gene ontology (GO) term enrichment using the TOPGO v.2.44.0 R package (Alexa & Rahenfuhrer, 2009). They were first grouped into four nonexclusive main categories depending on whether they were genetic differentiation outliers, or associated with temperature-related, precipitation-related or seasonality-related climate variables. Since GO term annotation for the *P. abies* genome is incomplete, we also adopted an *ad hoc* approach, specifically focusing on functions of interest, namely response to photoperiod, cold or detection of abiotic stimuli, growth, flowering and circadian clock.

We furthermore identified, with a newly developed approach, chromosome regions enriched for outlier SNPs (either low  $P$ -values in PCADAPT and LFMM2 analyses, or high Bayes factor in BAYENV2). The method (Tiret & Milesi, 2021) is described in the online methods and is freely accessible at <https://github.com/milesilab/peakdetection>. To do so, we used the consensus genetic map developed by Bernhardsson *et al.* (2019) with an overlapping set of probes and the same reference genome. We successfully placed on the Norway spruce consensus genetic map 89 940

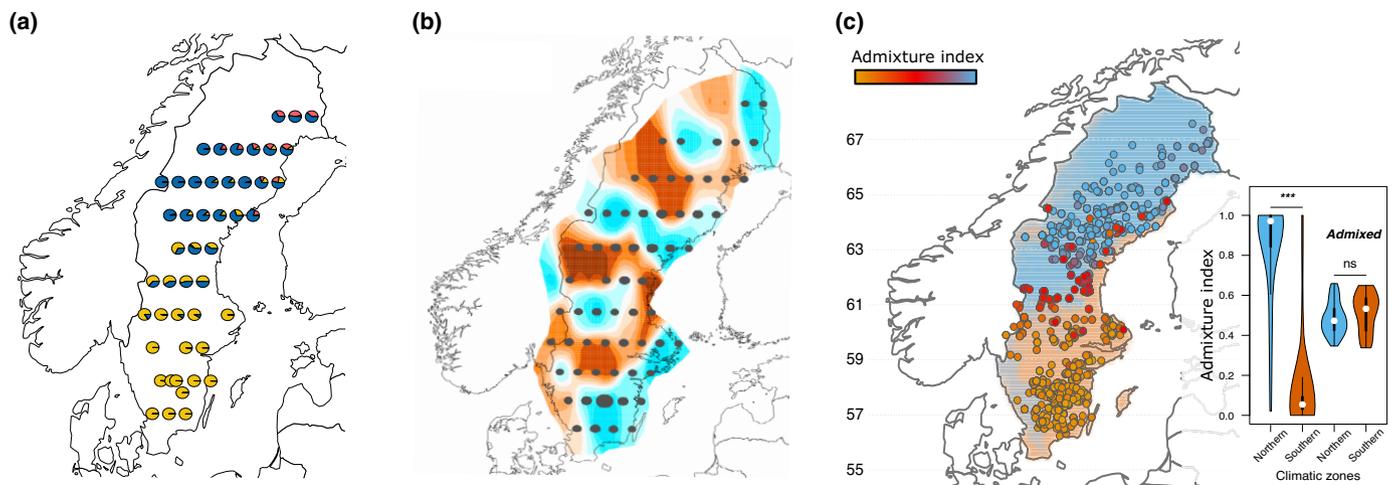
SNPs, captured with 16 559 probes, covering 15 137 different genes (*c.* 66% of the genes for which at least one SNP was called) and including *c.* 17% of the 28 354 ‘Well-supported genes’ described in Nystedt *et al.* (2013).

## Results

### Population structure and contact zone

We first investigated global population structure using the whole dataset ( $N=4769$ ), including *Picea obovata* samples. A limited number of *P. obovata* samples were used as an outgroup to help interpret *P. abies* genetic variation and population structure in a global context. We retrieved three main domains, Boreal, Carpathians and the Alps, and three additional clusters resulting from their admixture: Central Europe, Russia-Baltics and Northern Poland (Figs 1, S1, S2). Sweden is itself divided into two main genetics clusters, one including southern and central Sweden (hereafter, CSE) and the other one including the northern part of the country (hereafter, NFE). Many trees in southern Sweden also correspond to recent introductions from the rest of the Norway spruce natural range as historically documented (Myking *et al.*, 2016) and then genetically inferred (Chen *et al.*, 2019). Despite their current geographic closeness, the CSE and NFE clusters are divergent and CSE is more closely related to the Russia-Baltics cluster than to NFE ( $F_{\text{ST}}=0.009$  and  $0.018$ , respectively, Table S1). This general pattern is consistent with a recolonization of the Scandinavian peninsula from refugia with different genetic components and through two different routes, a Northern one and a Southern one. To study more finely the genetic structure of the contact zone and identify the evolutionary forces that shaped it, we focused in the rest of the study on the subset of trees that were native to Sweden and belonged to the CSE ( $N=974$ ) and the NFE ( $N=784$ ) clusters.

Regardless of the number of layers considered in CONSTRUCT analysis, a model including IBD within layers predicts the genetic variation pattern better than a nonspatial model (Fig. S3). The lowest cross-validation error (five-fold) was found for three layers (Fig. S3) but, in line with ADMIXTURE, two ancestry components explained most of the genetic variation and distinguished southern trees from northern ones (Fig. 1c). The contact zone between these two main clusters occurred between  $60^\circ\text{N}$  and  $63^\circ\text{N}$  (Fig. 2a). Contributions from the southern cluster into the northern one can be detected at latitudes as high as  $66^\circ\text{N}$  along the East coast while the northern cluster barely contributed to the populations outside of the contact zone. Finally, populations from high latitudes (close to  $67^\circ\text{N}$ ) also presented a specific ancestry component (Fig. 2a). Based on ADMIXTURE results this ancestry component probably represents more recent introgression from *P. obovata* into the northernmost *P. abies* populations (Fig. 1c), as supported by the large discrepancy in ancestry coefficients found between two Russian populations located at the same longitude (Fig. 1c). The effective migration surfaces estimated by EEMS result in a complex pattern, but regions with low effective migration rate correspond to the contact zone already detected by CONSTRUCT and to mountainous regions in the north.



**Fig. 2** Fine genetic structure of the contact zone and relation to climate zones in *Picea abies*. (a) ADMIXTURE proportions based on the best spatial model using CONSTRUCT ( $K = 3$ ). Colors represent different ancestry components. Close-by samples were grouped into 'populations'. (b) Estimated effective migration surfaces (EEMS). Blue and brown areas indicate regions with a higher or a lower effective migration rate than expected under a model of isolation by distance (IBD), respectively. Gray dots represent individual aggregations. (c) The genetic contact zone overlaps with the transition between the two main climatic zones, the southern one (orange background) and the northern one (blue background). Dots represent tree locations and the color scale corresponds to the admixture index (from 0, full CSE, orange, to 1, full NFE, blue). Violin plots represent the distribution of hybrid index within each of the two main climatic zones (all samples or only samples with  $0.33 > \text{hybrid index} < 0.66$ ; ns,  $P > 0.05$ ; \*\*\*,  $P < 0.001$ ). The white dot is the median and the black box the interval between the first and the third quartile.

North–South barriers, such as the one along the west coast, are probably artifacts due to the difficulty of EEMS to account for anisotropy (Petkova *et al.*, 2015) (Fig. 2b). Finally, considering all pairs of populations and using the framework developed by Rousset (1997), we detected a strong pattern of IBD, with a neighborhood size of  $N_s = 209 \pm 33$  individuals (Fig. S4). Gene flow was much less pronounced along a latitudinal gradient ( $N_s = 228 \pm 33$ ) than along a longitudinal gradient ( $N_s = 680 \pm 187$ ).

To investigate whether ecological barriers to gene flow contributed to the establishment of the contact zone, we analyzed environmental variation across Sweden. Using an unsupervised clustering approach, we delineated three climatic zones (Fig. 2c): the two main ones separate the northern part from the southern part of the country and the differentiation is mainly explained by temperature-related variables (annual mean temperature, minimum or average temperature of the coldest months, seasonality, Figs 2c, S5). The third climatic zone corresponds to the mountainous area and the west coast and is characterized by higher precipitation than the two other climatic zones. The genetic contact zone between the northern (NFE) and the southern (CSE) clusters almost perfectly overlaps the transition between the northern and southern climatic zones (Figs 2c, S5). Highly admixed trees between NFE and CSE ( $0.33 < A_{\text{ind}} \leq 0.66$ , see definition in the Materials and Methods section) were located along the transition zone and evenly distributed between the two climatic zones (Wilcoxon's rank-sum test,  $W = 789$ ,  $P = 0.18$ ; Fig. 2c), while most of the unadmixed trees belonging to the CSE cluster ( $A_{\text{ind}} \leq 0.33$ ) or to the NFE cluster ( $A_{\text{ind}} \geq 0.66$ ) were restricted to the southern or the northern climatic zone, respectively ( $W = 7 \times 10^{-5}$ ,  $P < 0.001$ ). Furthermore, there is a highly significant and quantitative relationship between the admixture index ( $A_{\text{ind}}$ ) and the bioclimatic variables that strongly discriminate the two main

climatic zones (Notes S1). While it could simply be the result of drift and migration, such a match between the main environmental zones and the genetic structure suggests that natural selection contributed to the creation and maintenance of the contact zone between the two genetics clusters.

### Genomic signatures of selection

We used coalescent simulations and an ABC framework to test for the presence of differential linked selection using DILS. We considered three datasets depending on the distance to the contact zone ('far', 'intermediate' and 'close'). For each dataset, the most likely model was the one with linked selection, with posterior probabilities of 71.24, 93.39 and 87.94%, respectively (Table 1). This suggests that linked selection occurred over the entire range of each climate zone. These results were further confirmed by forward simulations (Notes S2).

Genome scans identified 440 and 990 SNPs showing extreme allele frequency differences between geographic regions, using PCADAPT or  $\chi^2$  statistics respectively (32% overlap at the gene level). With GEA, a total of 1616 (BAYENV2) and 1298 (LFMM2) SNPs were associated with at least one of the bioclimatic variables (21% overlap at the gene level). The number of significant associations per bioclimatic variable was correlated between the two analyses (Spearman's  $\rho = 0.53$ ,  $S = 1070$ ,  $P < 0.01$ ) (Table S2). Most of the significant associations were with the climatic variables that contributed the most to the discrimination of the two main climatic zones (Spearman's  $\rho = 0.76$ ,  $S = 229.8$ ,  $P < 0.001$  and  $\rho = 0.65$ ,  $S = 350.15$ ,  $P < 0.01$ , respectively for LFMM2 and BAYENV2).

Candidate genes were significantly enriched for GO terms associated with biological processes related to environmental

**Table 1** Pairwise comparison of different models with Demographic Inference with Linked Selection (DILS) for individuals at varying distance from the center of the contact zone in Scandinavian *Picea abies*.

Distance to contact zone	AM vs SI	IM vs SC	M-homo vs M-hetero	$N_e$ -homo vs $N_e$ -hetero
Close	AM ( $P=1.00$ )	IM ( $P=0.52$ )	M-homo ( $P=0.89$ )	N-hetero ( $P=0.88$ )
Intermediate	AM ( $P=1.00$ )	IM ( $P=0.49$ )	M-homo ( $P=0.95$ )	N-hetero ( $P=0.93$ )
Far	AM ( $P=1.00$ )	IM ( $P=0.55$ )	M-homo ( $P=0.89$ )	N-hetero ( $P=0.71$ )

Demographic models: Strict Isolation (SI), Ancient Migration (AM), Isolation with Migration (IM), Secondary Contact (SC), Homogeneous and Heterogeneous migration ( $Nm$ ) (M-homo and M-hetero), and Homogeneous and Heterogeneous effective population size ( $N_e$ ) ( $N_e$ -homo and  $N_e$ -hetero). The value in parentheses,  $P$ , is the posterior probability of the best demographic model. Distance to contact zone ('close', 'intermediate', 'far') is defined according to the admixture index.

stimulus detection, metabolic pathways, growth and morphogenesis regulation, as well as biotic interactions (Fig. S6). Using an *ad hoc* approach specifically focusing on functions of interest (growth or phenology – related to or associated with detection of, or response to environmental variation), we identified 134 candidate SNPs located within or in the vicinity of 81 unique genes involved in these functions. We used a heatmap to illustrate how allele frequencies at these SNPs changed across populations (Fig. S7). Populations clustered according to latitude and this clustering was mostly driven by genes associated with the circadian clock and therefore to phenology and growth rhythm: *XAP5 time keeper* (Spearman's rho between allele frequencies and latitude = 0.70), *flowering-time-like loci* (*FTL*, rho = 0.59), *early flowering loci 3* (*EFL3*, rho = 0.92), *early flowering loci 3 high* (*EFL3-high*, rho = 0.91), *sensitivity to red light reduced 1* (*SSR1*, rho = 0.78) and *gigantea* (rho = 0.76) (Fig. 3a).

Genes putatively involved in local adaptation clustered in a limited number of genomic regions spread across the genome (four genes on average per region, maximum 14 for BAYENV2 analysis; six on average and maximum 22 for LFMM2), with one or several clusters on most linkage groups (Fig. 3b,c; Notes S1). All candidate regions with extreme allele frequency differences between geographic regions were associated with at least two environmental variables, suggesting a direct or indirect causal relationship between high genetic differentiation and environmental factors. Regions enriched for candidate genes were more often associated with temperature-related variables (on average  $4.5 \pm 3.6$  regions across the two GEA analyses, the maximum being nine for *temperature annual range*) than to precipitation-related ones ( $0.94 \pm 1.1$ , maximum being three for *precipitation of driest quarter*). The climatic variables that contributed the most to the discrimination of the two main climatic zones were also those for which we detected the highest number of genomic regions enriched for candidate genes (Spearman's rho = 0.65;  $S=469.32$ ;  $df=18$ ;  $P=0.002$  for BAYENV2). Similar results were obtained with LFMM2, the number of candidate genomic regions per variable being highly correlated between the two analyses

(rho = 0.63,  $S=853$ ,  $P<0.001$  and Fig. 3c). Genomic regions associated with local adaptation were found across all linkage groups but formed large clusters on individual chromosomes. Together, these results and those of the ABC analysis strongly support a significant contribution of natural selection to the establishment and maintenance of the contact zone.

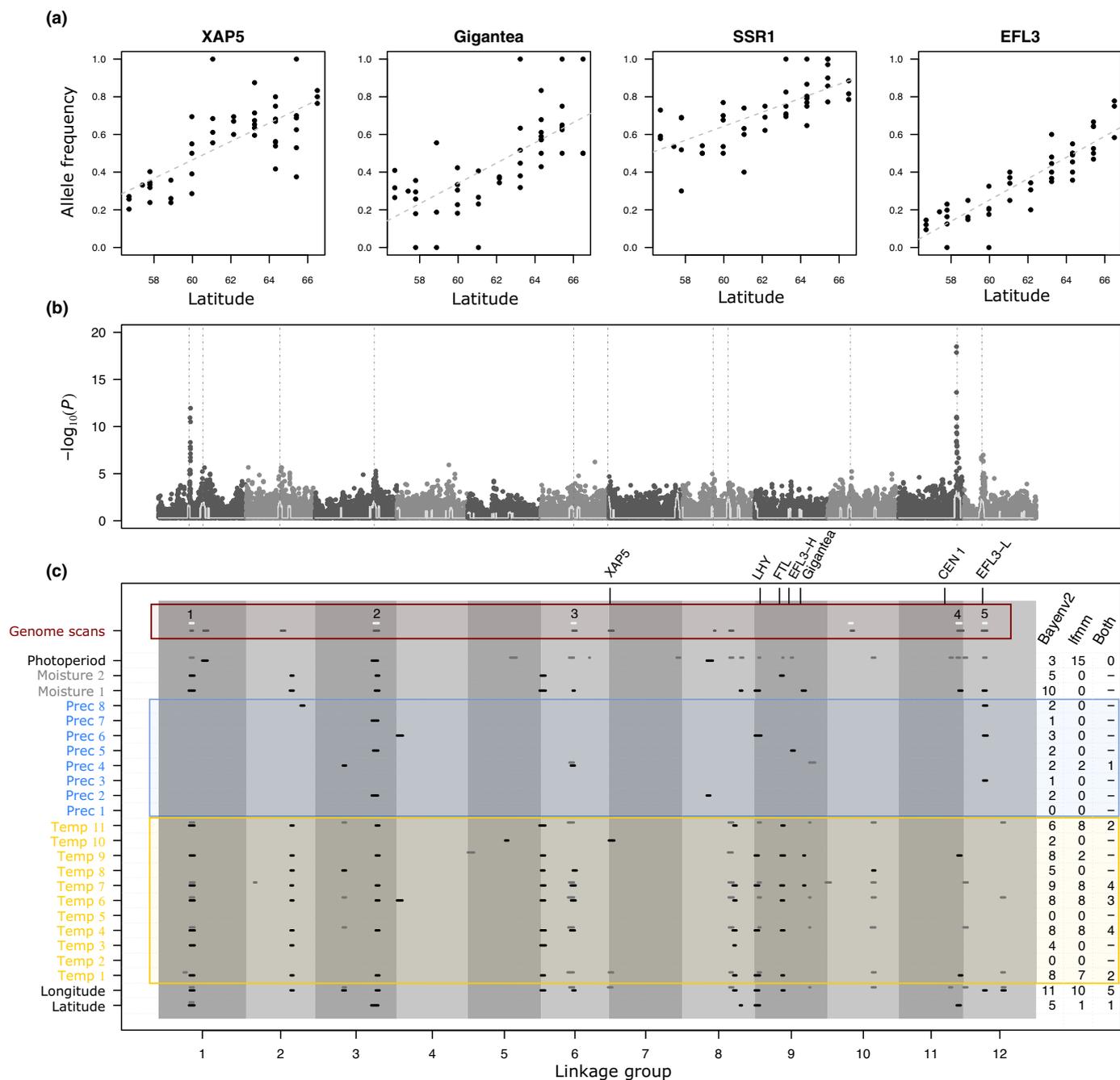
## Discussion

Contact zones are a rich source of information on the interplay between demography and selection in shaping the genetic structure of species (Johannesson *et al.*, 2020). In spite of their importance in shaping the genetic structure of most species, the origin and mechanism behind the maintenance of contact zones have received relatively little attention. Utilizing genomic data from almost 5000 trees sampled across Sweden and the natural range of Norway spruce, we reconstructed the origin of the contact zone separating the south and the north of Scandinavia and showed that the contact zone corresponds to a major climatic transition, and natural selection acting on gene clusters dispersed across the whole genome contributes to the maintenance of the differentiation between the two sides of the contact zone. Given that Norway spruce has been present in Scandinavia for a limited number of generations (Giesecke & Bennett, 2004; Nota *et al.*, 2022), this is an important result with respect to climate change because, unless trees were preadapted before invading Scandinavia, it suggests rapid local adaptation. It has often been assumed that trees, because of their long generation times, will be poorly equipped to respond rapidly to climate change (e.g. Dauphin *et al.*, 2021) but a recent study based on temporal change in allele frequencies in three oak stands from central France indicates that oaks responded readily to the climatic conditions of the Little Ice Age as well as, later on, to the warming period (Saleh *et al.*, 2022). Interestingly, as in the present case, the response to selection was highly polygenic and genome-wide.

### A recent contact zone

The general clustering is congruent with what was observed in previous studies using smaller sample sizes (Chen *et al.*, 2019) or different markers (Tsuda *et al.*, 2016). According to these population genetics studies and the paleoecological record (pollen fossil data but also macrofossils) (Giesecke & Bennett, 2004; Latałowa & van der Knaap, 2006; Binney *et al.*, 2009, 2017; Lehsten *et al.*, 2014), current European populations of *P. abies* originated from at least three main ancient refugia located in the Alps, in the Carpathians, and in the Russian and Western Siberia Plains. What our data show is that these three lineages did not evolve independently but rather entered into contact at many points in response to cyclic variation in climate. For example, as apparent from the ADMIXTURE analysis, both Northern Poland and the Russian-Baltic domain are three-way admixture zones, with a major contribution from the Carpathians and more limited contributions from the Alps and *P. obovata*.

The recolonization of Northern Europe by *P. abies* started relatively late, and spruce migration rates for Fennoscandia varied



**Fig. 3** Genome scans for differentiation outliers and genotype–environment association (GEA) in *Picea abies*. (a) Examples of allele frequency variation as a function of latitude for four candidate genes involved in control of the circadian clock. (b) Manhattan plot ( $-\log_{10} P$ -values) of genome scans for an excess of differentiation (PCADAPT). Dark and light gray backgrounds delineate linkage groups. Vertical dotted gray lines represent regions enriched for low  $P$ -values, ‘peaks’ in the profile. A detailed analysis is provided for each of the genome scans and GEA in Notes S1. (c) For each genome scan (white,  $\times^T$ ; gray, PCADAPT) and GEA (black, BAYENV2; dark gray, LFMM2), significant peaks are localized on the Norway spruce genetic map. For each geographic and bioclimatic variable, the number of significant peaks is indicated on the right as well as the number of shared peaks. Numbers at the top of the graph identify significant peaks detected by the two genome scan methods and at least one GEA method. When possible, genes involved in control of the circadian clock were placed onto the genetic map.

between 200 and 500 m yr<sup>-1</sup> (Lehsten *et al.*, 2014). As for many other species, for example humans (Günther *et al.*, 2018; Peter *et al.*, 2020), aspens (De Carvalho *et al.*, 2010), brown bears (Bray *et al.*, 2013) and rodents (Jaarola *et al.*, 1999), our data support the existence of two routes of recolonization of Scandinavia, both from east to west, but one entering Scandinavia from

the north and moving southward and one entering Scandinavia at a lower latitude and moving both northward and southward (Giesecke & Bennett, 2004; Nota *et al.*, 2022). The two routes joined between 60°N and 63°N and created an admixture zone that was identified in the present study. Fossil data indicate that trees entered Scandinavia around 13 000–12 000 yr ago from the

south and 4000–3000 yr ago from the north (Latałowa & van der Knaap, 2006). The recolonization of Scandinavia by Norway spruce occurred in two phases: a first phase during which small outposts were established and, later on, a second phase when dispersal from those and from a larger front started (Latałowa & van der Knaap, 2006). If their average migration rate was  $300 \text{ m yr}^{-1}$ , trees should have reached the current location of the contact zone after around 3300 and 2000 yr ago, respectively. So, the contact zone would have been created some 2000 yr ago, or, assuming a generation time of around 50 yr, some 40 generations ago. The pollen fossil record suggests a somewhat lower migration rate and the fronts reaching central Sweden some 3000 yr ago, so around 60 generations ago. Of course, these are approximate dates and we do not expect the northwards and southwards migrations to progress at similar speed as it is a well-established fact that Norway spruce can easily be transferred some 3–4 degrees of latitude northwards without much loss in growth but that a southwards move is generally much less successful (Eriksson & Ekberg, 2001). We indeed observed an asymmetry, with the southern cluster contributing to the northern one at a latitude as high as  $66^\circ\text{N}$  while the northern cluster contribution to the southern one was much more limited. Regardless, given that gene flow is important in Norway spruce, this implies that we would have expected the contact zone to have started to be eroded by gene flow unless it were maintained by selection.

In addition to the main contact zone, in the CONSTRUCT analysis, northernmost populations contain an ancestral component that was specific to those populations (Fig. 2a, red component). A similar result was obtained by Androsiuk *et al.* (2013) that showed, based on 15 SRR gene loci, that breeding populations from the northernmost range of Sweden clustered separately and presented signs of a bottleneck. Those populations are also characterized by a higher contribution from *P. obovata* (Tollefsrud *et al.*, 2009). Our study provides further support for asymmetrical introgression from *P. obovata* into the western range of *P. abies* with a larger contribution at high latitudes (*c.*  $65^\circ\text{N}$  and above) than at intermediate ones (*c.*  $60^\circ\text{N}$ ) (Tsuda *et al.*, 2016; Fagernäs, 2017; Chen *et al.*, 2019). This pattern at high latitudes is not specific to the *P. abies*–*P. obovata* species pair. A similar situation is observed between *Larix sibirica* and *L. gmelinii* with introgression of mitochondrial DNA from the local species in the west, *L. sibirica*, into the invading species from the East, *L. gmelinii* (Polezhaeva *et al.*, 2010; Semerikov *et al.*, 2013; Schulte *et al.*, 2021). This trend does not preclude migration in the opposite direction. For example, *Pinus sylvestris* apparently dispersed primarily from western Europe (Semerikov *et al.*, 2020).

Finally, pollen analysis, genetic data and simulations supported a moving front of recolonization of Scandinavia (Giesecke & Bennett, 2004; Giesecke, 2005; Lehsten *et al.*, 2014; Nota *et al.*, 2022) rather than population expansion from local refugia (Parducci *et al.*, 2012). Putative local refugia have been found in mountainous areas of central Sweden (Kullman, 1996) and might have had a local impact, but the fit to an IBD pattern, together with the importance of the contribution of *P. obovata*, would rather argue for a recolonization from populations located outside of the main glaciated areas. Also, these refugial populations

consist of small trees that reproduce mainly asexually (Kullman, 1996) and it is unlikely, even if they started to reproduce sexually and expanded once the climate became warmer, that they could have contributed massively to surrounding populations. More generally, comparison between *Picea* and *Larix* in eastern Siberia suggests that the biology of *Picea* (relatively heavy seeds, low genetic diversity in survival pockets) might explain why *Larix* and not *Picea* was capable of population expansion from small, scattered refugia (Herzschuh, 2020).

### Polygenic architecture of local adaptation along the contact zone and ecological drivers

We have so far discussed the data in terms of demographic events. However, the major contact zone observed in Scandinavia corresponds to a discontinuity in bioclimatic factors, is better explained by a model incorporating linked selection than by a purely neutral one and is accompanied by a large number of genomic areas containing clusters of genes characterized by high genetic differentiation and association with climatic variables related to latitude (e.g. photoperiod, temperature-related climatic variables). While this is not the first study of forest trees to indicate the presence of selection along a latitudinal gradient, it is the first to demonstrate the genome-wide impact of local adaptation. The observed genomic pattern is expected under polygenic adaptation for different optima when populations are linked by gene flow (Yeaman & Whitlock, 2011; Yeaman, 2013) and could be further reinforced or even caused by structural rearrangements that allow the spread of coadapted alleles. Unfortunately, the current state of the *P. abies* genome assembly (20 Gb, >15 million scaffolds) does not enable us to investigate this hypothesis further. However, as the largest region includes up to 22 genes carried by different scaffolds, we can expect that some regions enriched for candidate genes are structural variants that can further limit gene flow between the northern and southern clusters.

A large number of genes were significantly associated with environmental variables and were differentiation outliers: 205 unique genes carried at least one significant SNP associated with environmental variables and 91 were outliers in genome scans. Among the latter, some may be false positives as the populations are recent and not at demographic equilibrium (Bierne *et al.*, 2013; Lotterhos & Whitlock, 2014). However, because of the confounded effects of population structure and of the main environmental gradient, these numbers are likely to be underestimated (Milesi *et al.*, 2019), but this result, nevertheless, suggests a high degree of polygenicity of local adaptation in Norway spruce, in line with Milesi *et al.* (2019) and Chen *et al.* (2021). Notably, the pattern of IBD was stronger at the identified candidate genes than at all genes considered together. Many candidate genes were involved in circadian clock control (*XAP5*, *FTL*, *EFL-3*, *EFL-3 high*, *Gigantea*, *CEN1*, *SRR1*, *LHY*), together suggesting that selection on phenology could partly explain the maintenance of the contact zone in Norway spruce by inducing differences in reproductive period and limiting gene flow. Interestingly, three important genes for phenology, *FTL*, *EFL-3* and *Gigantea*, are located nearby on linkage group 8 (Fig. 3c). This colocalization

could be advantageous in the context of strong selection pressure, such as the one exerted on juvenile trees by frost in late spring and early fall (Hannerz, 1998), as it allows for the cosegregation of adaptive loci involved in the control of phenology.

Considering the overall low population genetic differentiation together with the relatively short time spent by trees in Scandinavia, the establishment of such a strong clinal gradient would seem to imply a rather strong selection pressure, even at individual loci. Assuming that local refugia did not contribute significantly to the recolonization of Scandinavia, Norway spruce entered Scandinavia around 10 000–12 000 cal yr BP and reached central Sweden around 3000 cal yr BP (Latalowa & van der Knaap, 2006; Tollefsrud *et al.*, 2009), and considering a generation time of about 50 yr implies that the observed gradient at adaptive loci over Sweden was established in around 150–200 generations. However, it cannot be ruled out that preadapted loci also contributed to local adaptation in newly colonized areas. As trees from the two main clusters originated from similar latitudes to those found today in Scandinavia, a certain level of preadaptation seems likely. Additional samples from northwestern Russia and from the Baltics would be necessary to test this hypothesis.

### Predicting the response to climate change of the contact zone and practical implications

Our results indicate that the current contact zone is maintained by natural selection and will therefore change as the climate changes. Three main scenarios for the response of Scandinavian populations under rapid climate change seem plausible. First, trees from the northern cluster (NFE) are progressively going to be introgressed with genes from the southern cluster (CSE) as the latter moves northwards and the contact zone progressively disappears. Second, barriers to gene flow are strong enough between the two clusters for the contact zone to persist and shift northwards. Third, assuming that growth traits are a good proxy for fitness, global change will be advantageous for populations with more southern origins, for instance favoring trees with an Alpine or Carpathian genetic background and those will progressively replace existing populations. Given that Milesi *et al.* (2019) showed that, at least in the southern and central parts of Sweden, trees with an Alpine or a Carpathian origin outperformed the trees from local provenance for growth traits, this may well occur.

The response of the Scandinavian population of Norway spruce to climate change will strongly impact breeding, especially if breeding is based on genomic selection (Meuwissen *et al.*, 2001). Genomic selection is increasingly being adopted in forest tree breeding programs (Grattapaglia *et al.*, 2018), but it lacks accuracy when the training set is grown in an environment that differs from the target environment (Resende Jr. *et al.*, 2012). The evolution of the contact zone will thus need to be monitored and incorporated into future genotype-by-climate zone interaction studies to optimize the delineation of breeding zones, something that, to the best of our knowledge, has not yet been implemented in forest tree breeding. Indeed, the prediction of genotype by environment interactions in an unobserved

environment is today still challenging, so that genomic prediction is preferentially implemented with at least one training set per breeding zone in forest trees (Grattapaglia, 2017).

Predicting the future evolution of natural populations given global change, for instance for conservation and breeding, is and will remain a complex task, even more so for species such as Norway spruce that are tightly associated with human activity. The detection of adaptive loci that are associated with phenotypic traits and/or the environment will not be sufficient to predict future adaptation under climate change scenarios without in-depth knowledge of both global and local genetic diversity and how this diversity translates into fitness under various environments. Indeed, adaptation to a highly dimensional environment requires a high degree of polygenicity. It is therefore intrinsically challenging to extrapolate both genotype–phenotype and genotype–environment relationships under various scenarios involving either demographic or environmental changes. A possible approach would be to consider monitoring the delineation of breeding zones, and performing association analyses such as the one carried out in the present paper in each of them. The robustness of the genomic prediction will then no longer be a matter of genotype by environment interactions, but of the predictive power of future environments (Grattapaglia, 2017). The final challenge would be to incorporate introgression from closely related species (or from individuals from outside the focal range), and account for its role in the prediction in shaping genetic diversity and response to the environment. Predicting the phenotypes of crossbred populations is a challenging task and requires the implementation of special training sets (e.g. Misztal *et al.*, 2020), therefore warranting further studies to adapt these methods to natural population studies.

### Acknowledgements

We are grateful to Luis Leal and Linus Söderquist for comments on early versions of the manuscript and to Camille Roux for help with running DILS. This project is supported by the Swedish Foundation for Strategic Research (SSF), grant no. RBP14-0040. MT was supported by the EU H2020 project B4EST. Computation and data handling were provided by the Swedish National Infrastructure for Computing (SNIC) at Uppmax, partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

### Author contributions

ML and PM designed the study; LL, PM, MT, JC and JS analyzed the data; JB, ZC, LZ, BK, MB, JW, RGG and HW performed experiments or field work and contributed data; LL, PM, MT and ML drafted the manuscript. All authors read and approved the final version of the manuscript. LL and PM contributed equally to this work.

### ORCID

Mats Berlin  <https://orcid.org/0000-0002-7099-3322>

Jun Chen  <https://orcid.org/0000-0003-3187-7668>  
 Zhi-qiang Chen  <https://orcid.org/0000-0001-9725-8929>  
 Maria Rosario Garcia-Gil  <https://orcid.org/0000-0002-6834-6708>  
 Martin Lascoux  <https://orcid.org/0000-0003-1699-9042>  
 Pascal Milesi  <https://orcid.org/0000-0001-8580-4291>  
 Mathieu Tiret  <https://orcid.org/0000-0001-5389-4886>  
 Harry X. Wu  <https://orcid.org/0000-0002-7072-4704>

## Data availability

The data that support the findings of this study are openly available in NCBI, [PRJNA511374](https://doi.org/10.1111/nph.18480) and [PRJNA731384](https://doi.org/10.1111/nph.18480).

## References

- Alberto FJ, Aitken SN, Alía R, González-Martínez SC, Hänninen H, Kremer A, Lefevre F, Lenormand T, Yeaman S, Whetten R *et al.* 2013. Potential for evolutionary responses to climate change: evidence from tree populations. *Global Change Biology* 19: 1645–1661.
- Alexa A, Rahenfuhrer J. 2009. TOPGO: enrichment analysis for gene ontology. *Bioconductor* 27: 1–26.
- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19: 1655–1664.
- Androsiuk P, Shimono A, Westin J, Lindgren D, Fries A, Wang X-R. 2013. Genetic status of Norway spruce (*Picea abies*) breeding populations for northern Sweden. *Silvae Genetica* 62: 127–136.
- Bennett KD. 1997. *Evolution and ecology: the pace of life*. Cambridge, UK: Cambridge University Press.
- Bernhardsson C, Vidalis A, Wang X, Scofield DG, Schiffthaler B, Baison J, Street NR, García-Gil MR, Ingvarsson PK. 2019. An Ultra-dense haploid genetic map for evaluating the highly fragmented genome assembly of Norway Spruce (*Picea abies*). *G3* 9: 1623–1632.
- Bierne N, Roze D, Welch JJ. 2013. Pervasive selection or is it...? why are  $F_{ST}$  outliers sometimes so frequent? *Molecular Ecology* 22: 2061–2064.
- Binney H, Edwards M, Macias-Fauria M, Lozhkin A, Anderson P, Kaplan JO, Andreev A, Bezrukova E, Blyakharchuk T, Jankovska V *et al.* 2017. Vegetation of Eurasia from the last glacial maximum to present: key biogeographic patterns. *Quaternary Science Reviews* 157: 80–97.
- Binney HA, Willis KJ, Edwards ME, Bhagwat SA, Anderson PM, Andreev AA, Blaauw M, Dambon F, Haesaerts P, Kienast F *et al.* 2009. The distribution of late-Quaternary woody taxa in northern Eurasia: evidence from a new macrofossil database. *Quaternary Science Reviews* 28: 2445–2464.
- Bradburd GS, Coop GM, Ralph PL. 2018. Inferring continuous and discrete population genetic structure across space. *Genetics* 210: 33–52.
- Bray SCE, Austin JJ, Metcalf JL, Østbye K, Østbye E, Lauritzen S, Aaris-Sørensen K, Valdiosera C, Adler CJ, Cooper A. 2013. Ancient DNA identifies post-glacial recolonisation, not recent bottlenecks, as the primary driver of contemporary mtDNA phylogeography and diversity in Scandinavian brown bears. *Diversity and Distributions* 19: 245–256.
- Bucci G, Vendramin GG. 2000. Delineation of genetic zones in the European Norway spruce natural range: preliminary evidence. *Molecular Ecology* 9: 923–934.
- Caye K, Jumentier B, Lepeule J, François O. 2019. LFMM2: Fast and accurate inference of gene-environment associations in genome-wide studies. *Molecular Biology and Evolution* 36: 852–860.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4: 1–16.
- Chen J, Kallman T, Ma X, Gyllenstrand N, Zaina G, Morgante M, Bousquet J, Eckert A, Wegryzn J, Neale DB *et al.* 2012. Disentangling the roles of history and local selection in shaping clinal variation of allele frequencies and gene expression in Norway Spruce (*Picea abies*). *Genetics* 191: 865–881.
- Chen J, Li L, Milesi P, Jansson G, Berlin M, Karlsson B, Aleksic J, Vendramin GG, Lascoux M. 2019. Genomic data provide new insights on the demographic history and the extent of recent material transfers in Norway spruce. *Evolutionary Applications* 12: 1539–1551.
- Chen J, Tsuda Y, Stocks M, Kallman T, Xu N, Karkkainen K, Huotari T, Semerikov VL, Vendramin GG, Lascoux M. 2014. Clinal variation at phenology-related genes in spruce: parallel evolution in FTL2 and gigantea? *Genetics* 197: 1025–1038.
- Chen Z-Q, Zan Y, Milesi P, Zhou L, Chen J, Li L, Cui B, Niu S, Westin J, Karlsson B *et al.* 2021. Leveraging breeding programs and genomic data in Norway spruce (*Picea abies* L. Karst) for GWAS analysis. *Genome Biology* 22: 179.
- Coop G, Witonsky D, Di Rienzo A, Pritchard JK. 2010. Using environmental correlations to identify loci underlying local adaptation. *Genetics* 185: 1411–1423.
- Dauphin B, Rellstab C, Schmid M, Zoller S, Karger DN, Brodbeck S, Guillaume F, Gugerli F. 2021. Genomic vulnerability to rapid climate warming in a tree species with a long generation time. *Global Change Biology* 27: 1181–1195.
- De Carvalho D, Ingvarsson PK, Joseph J, Suter L, Sedivy C, Macaya-Sanz D, Cottrell J, Heinze B, Schanzer I, Lexer C. 2010. Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree. *Molecular Ecology* 19: 1638–1650.
- Diaz-Papkovich A, Anderson-Trocme L, Ben-Eghan C, Gravel S. 2019. UMAP reveals cryptic population structure and phenotype heterogeneity in large genomic cohorts. *PLoS Genetics* 15: e1008432.
- Eriksson G, Ekberg I. 2001. *An introduction to forest genetics*. Uppsala, Sweden: Swedish University of Agricultural Sciences.
- Fagnäs Z. 2017. *Biogeography of Norway spruce (Picea abies (L.) Karst.): Insights from a genome-wide study*, MSc thesis. Umeå University, Umeå, Sweden.
- Fraïsse C, Popovic I, Mazoyer C, Spataro B, Delmotte S, Romiguier J, Loire É, Simon A, Galtier N, Duret L *et al.* 2021. DILS: Demographic inferences with linked selection by using ABC. *Molecular Ecology Resources* 21: 2629–2644.
- Galinsky KJ, Bhatia G, Loh P-R, Georgiev S, Mukherjee S, Patterson NJ, Price AL. 2016. Fast principal-component analysis reveals convergent evolution of ADH1B in Europe and East Asia. *The American Journal of Human Genetics* 98: 456–472.
- Giesecke T. 2005. Holocene dynamics of the Southern boreal forest in Sweden. *The Holocene* 15: 858–872.
- Giesecke T, Bennett KD. 2004. The Holocene spread of *Picea Abies* (L) Karst in Fennoscandia and adjacent areas. *Journal of Biogeography* 31: 1523–1548.
- Grattapaglia D. 2017. Status and perspectives of genomic selection in forest tree breeding. In: *Genomic selection for crop improvement*. Cham, Switzerland: Springer, 199–249.
- Grattapaglia D, Silva-Junior OB, Resende RT, Cappa EP, Müller BSF, Tan B, Isik F, Ratcliffe B, El-Kassaby YA. 2018. Quantitative genetics and genomics converge to accelerate forest tree breeding. *Frontiers in Plant Science* 9: 1693.
- Günther T, Coop G. 2013. Robust identification of local adaptation from allele frequencies. *Genetics* 196: 205–220.
- Günther T, Malmström H, Svensson EM, Omrak A, Sánchez-Quinto F, Kilinc GM, Krzewińska M, Eriksson G, Fraser M, Edlund H *et al.* 2018. Population genomics of Mesolithic Scandinavia: investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biology* 16: e2003703.
- Hannerz M. 1998. *Genetic and seasonal variation in hardiness and growth rhythm in boreal and temperate conifers—a review and annotated bibliography*. Report no.2 Uppsala, Sweden: Skogforsk.
- Herzschuh U. 2020. Legacy of the Last glacial on the present-day distribution of deciduous versus evergreen boreal forests. *Global Ecology and Biogeography* 29: 198–206.
- Heuertz M, Paoli ED, Kallman T, Larsson H, Jurman I, Morgante M, Lascoux M, Gyllenstrand N. 2006. Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. *Genetics* 174: 2095–2105.
- Hewitt G. 2000. The genetic legacy of the quaternary ice ages. *Nature* 405: 907–913.
- Hudson RR, Slatkin M, Maddison WP. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* 132: 583–589.
- Jaarola M, Tegelström H, Fredga K. 1999. Colonization history in Fennoscandian rodents. *Biological Journal of the Linnean Society* 68: 113–127.

- Johannesson K, Moan AL, Perini S, André C. 2020. A Darwinian laboratory of multiple contact zones. *Trends in Ecology & Evolution* 35: 1021–1036.
- Kullman L. 1996. Norway Spruce present in the Scandes Mountains, Sweden at 8000 BP: new light on Holocene tree spread. *Global Ecology and Biogeography Letters* 5: 94–101.
- Lagercrantz U, Ryman N. 1990. Genetic structure of Norway spruce (*Picea abies*): concordance of morphological and allozymic variation. *Evolution* 44: 38–53.
- Latalowa M, van der Knaap WO. 2006. Late Quaternary expansion of Norway spruce *Picea abies* (L.) Karst in Europe according to pollen data. *Quaternary Science Reviews* 25: 2780–2805.
- Lehsten D, Dullinger S, Hülber K, Schurgers G, Cheddadi R, Laborde H, Lehsten V, François L, Dury M, Sykes MT. 2014. Modelling the Holocene migrational dynamics of *Fagus sylvatica* L. and *Picea abies* (L.) H. Karst. *Global Ecology and Biogeography* 23: 658–668.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup 1000 Genome Project Data Processing. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078–2079.
- Lotterhos KE, Whitlock MC. 2014. Evaluation of demographic history and neutral parameterization on the performance of  $F_{ST}$  outlier tests. *Molecular Ecology* 23: 2178–2192.
- Luu K, Bazin E, Blum MG. 2017. PCADAPT: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources* 17: 67–77.
- Meuwissen THE, Hayes BJ, Goddard M. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.
- Milesi P, Berlin M, Chen J, Orsucci M, Li L, Jansson G, Karlsson B, Lascoux M. 2019. Assessing the potential for assisted gene flow using past introduction of Norway spruce in southern Sweden: local adaptation and genetic basis of quantitative traits in trees. *Evolutionary Applications* 12: 1946–1959.
- Misztal I, Lourenco D, Legarra A. 2020. Current status of genomic evaluation. *Journal of Animal Science* 98: skaa101.
- Moran BM, Payne C, Langdon Q, Powell DL, Brandvain Y, Schumer M. 2021. The genomic consequences of hybridization. *eLife* 10: e69016.
- Myking T, Rusanen M, Steffenrem A, Kjær ED, Jansson G. 2016. Historic transfer of forest reproductive material in the Nordic region: drivers, scale and implications. *Forestry* 89: 325–337.
- Nota K, Klaminder J, Milesi P, Bindler R, Nobile A, van Steijn T, Bertilsson T, Svensson B, Hirota SK, Matsuo A *et al.* 2022. Norway spruce postglacial recolonization of Fennoscandia. *Nature Communications* 13: 1333.
- Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A *et al.* 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature* 497: 579–584.
- Parducci L, Jorgensen TJ, Tollefsrud MM, Elverland E, Alm T, Fontana SL, Bennett KD, Haile J, Matetovici I, Suyama Y *et al.* 2012. Glacial survival of boreal trees in northern Scandinavia. *Science* 335: 1083–1086.
- Peter BM, Petkova D, Novembre J. 2020. Genetic landscapes reveal how human genetic diversity aligns with geography. *Molecular Biology and Evolution* 37: 943–951.
- Petkova D, Novembre J, Stephens M. 2015. Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics* 48: 94–100.
- Polezhaeva MA, Lascoux M, Semerikov VL. 2010. Cytoplasmic DNA variation and biogeography of *Larix Mill* in Northeast Asia. *Molecular Ecology* 19: 1239–1252.
- Privé F, Luu K, Vilhjálmsson BJ, Blum MG. 2020. Performing highly efficient genome scans for local adaptation with R Package PCADAPT v.4. *Molecular Biology and Evolution* 37: 2153–2154.
- Resende MFR Jr, Muñoz P, Acosta JJ, Peter GF, Davis JM, Grattapaglia D, Resende MDV, Kirst M. 2012. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytologist* 193: 617–624.
- Rousset F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145: 1219–1228.
- Saleh D, Chen J, Leplé J, Leroy T, Truffaut L, Dencausse B, Lalanne C, Labadie K, Lesur I, Bert D *et al.* 2022. Genome-wide evolutionary response of European oaks during the Anthropocene. *Evolution Letters* 6: 4–20.
- Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation. *Nature Reviews Genetics* 14: 807–820.
- Savolainen O, Pyhäjärvi T, Knürr T. 2007. Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics* 38: 595–619.
- Schulte L, Bernhardt N, Stoof-Leichsenring K, Zimmermann HH, Pestryakova LA, Epp LS, Herzsich U. 2021. Hybridization capture of larch (*Larix Mill*) chloroplast genomes from sedimentary ancient DNA reveals past changes of Siberian forest. *Molecular Ecology Resources* 21: 801–815.
- Semerikov NV, Petrova IV, Sannikov SN, Semerikova SA, Tashev AN, Lascoux M, Semerikov VL. 2020. Cytoplasmic DNA variation does not support a recent contribution of *Pinus sylvestris* L. from the Caucasus to the main range. *Tree Genetics & Genomes* 16: 59.
- Semerikov VL, Semerikova SA, Polezhaeva MA, Kosintsev PA, Lascoux M. 2013. Southern montane populations did not contribute to the recolonization of West Siberian Plain by Siberian larch (*Larix sibirica*): a range-wide analysis of cytoplasmic markers. *Molecular Ecology* 22: 4958–4971.
- Tiret M, Milesi P. 2021. Statistical peak detection for GWAS. *Zenodo*. [WWW document] URL <https://doi.org/10.5281/zenodo.5079556> [accessed 7 July 2021].
- Tollefsrud MM, Kissling R, Gugerli F, Johnsen O, Skroppa T, Cheddadi R, van der Knaap WO, Latalowa M, Terhürne-Berson RT, Litt T *et al.* 2008. Genetic consequences of glacial survival and postglacial colonization in Norway spruce: combined analysis of mitochondrial DNA and fossil pollen. *Molecular Ecology* 17: 4134–4150.
- Tollefsrud MM, Latalowa M, van der Knaap WO, Brochmann C, Sperisen C. 2015. Late Quaternary history of North Eurasian Norway spruce (*Picea abies*) and Siberian spruce (*Picea obovata*) inferred from macrofossils, pollen and cytoplasmic DNA. *Journal of Biogeography* 42: 1431–1442.
- Tollefsrud MM, Sonstebo JH, Brochmann C, Johnsen O, Skroppa T, Vendramin GG. 2009. Combined analysis of nuclear and mitochondrial markers provide new insight into the genetic structure of North European *Picea abies*. *Heredity* 102: 549–562.
- Tsuda Y, Chen J, Stocks M, Kallman T, Sonstebo JH, Parducci L, Semerikov V, Sperisen C, Politov D, Ronkainen T *et al.* 2016. The extent and meaning of hybridization and introgression between Siberian spruce (*Picea obovata*) and Norway spruce (*P. abies*): cryptic refugia as stepping stones to the west? *Molecular Ecology* 25: 2773–2789.
- Vendramin GGG, Anzidei MM, Madaghiale AA, Sperisen CC, Bucci GG. 2000. Chloroplast microsatellite analysis reveals the presence of population subdivision in Norway spruce (*Picea abies* K.). *Genome* 43: 68–78.
- Vidalis A, Scofield DG, Neves LG, Bernhardsson C, García-Gil MR, Ingvarsson PK. 2018. Design and evaluation of a large sequence-capture probe set and associated SNPs for diploid and haploid samples of Norway spruce (*Picea abies*). *BioRxiv*. doi: [10.1101/291716](https://doi.org/10.1101/291716).
- Wang X, Bernhardsson C, Ingvarsson PK. 2020. Demography and natural selection have shaped genetic variation in the widely distributed conifer Norway spruce (*Picea abies*). *Genome Biology and Evolution* 12: 3803–3817.
- Wielstra B. 2019. Historical hybrid zone movement: more pervasive than appreciated. *Journal of Biogeography* 46: 1300–1305.
- Wielstra B. 2021. Hybrid zones. *Current Biology* 31: R108–R109.
- Yeaman S. 2013. Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences, USA* 110: E1743–E1751.
- Yeaman S, Whitlock MC. 2011. The genetic architecture of adaptation under migration–selection balance. *Evolution* 65: 1897–1911.

## Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

**Fig. S1** Principal components analysis based on resampling of 80 samples per cluster.

**Fig. S2** Admixture plot ( $K=2$  to 7) and cross-validation error plot.

**Fig. S3** Construct cross-validation and layer contribution plots.

**Fig. S4** Isolation by distance pattern for all single nucleotide polymorphisms (SNPs) or candidate SNPs.

**Fig. S5** Relative contribution of the climatic variable to each of the three climate zones.

**Fig. S6** Scatter plot based on GO term enrichment analysis.

**Fig. S7** Heatmap of allele frequency variation between population for candidate single nucleotide polymorphisms detected with the *ad hoc* procedure.

**Methods S1** Additional online methods.

**Notes S1** Relationship between climatic zones and Admixed index.

**Notes S2** Simulating a neutral demography of NFE vs CSE with forward simulations.

**Table S1** Pairwise  $F_{ST}$  index estimates among *Picea obovata*, two admixed *P. abies* × *P. obovate* populations (Indigo and Kirov), and the seven main *P. abies* genetic clusters.

**Table S2** Number of candidate loci identified through genotype–environment associations.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.