BACKGROUND

Human activities have led to the destruction of the habitat of an increasing number of species and climate change has only exacerbated this situation (Bellard et al., 2012; Grimm et al., 2013; Walther et al., 2002). Given the rapid pace of global warming, ecological communities face an urgent need for rapid adaptation in their structure and composition to cope with these ongoing challenges. Interestingly, recent studies have revealed that communities tend to respond more strongly to regional than global changes (Altieri & Gedan, 2015; Araújo & Rahbek, 2006; Atkins & Travis, 2010; Thuiller et al., 2005; Williams et al., 2021; Wuethrich, 2000). Climate change impacts vary across ecosystems, leading to diverse ecological dynamics and unpredictable changes within local environments (Carbeck et al., 2022; Walther et al., 2002). In order to mitigate the risk of demographic collapse and potential extinction, organisms must either track their climatic niche, be phenotypically plastic, or adapt to the changing environment from existing genetic variation.
or through de novo mutations. Adaptations of populations to local conditions result from a subtle balance between the intensity of natural selection and migration between the populations. While the former pushes the populations toward their local optima, the latter tends to homogenize the gene pools between the populations (for a review, see Bonnet et al., 2022; Lenormand, 2002). Still, some level of gene flow, although not an absolute requirement, contributes to promoting local and global genetic diversity and preserving the adaptive potential of populations (e.g., Frankham, 2005; Markert et al., 2010; Pauls et al., 2013). Global climate change poses significant challenges as it can affect simultaneously many populations, placing species and their populations at greater vulnerability in limiting their ability to cope with environmental changes and potentially leading to extinction (Leroy et al., 2020; Pauls et al., 2013; Rellstab et al., 2016).

Plants face heightened vulnerability to global change compared to animals due to challenges in tracking suitable habitats, suggesting species with longer generation times may encounter increased difficulty in adapting to environmental shifts (Dauphin et al., 2021; Leroy et al., 2020). Their survival hinges on the dual capacity to expand into new regions, drawing on past adaptations, and for edge populations to adapt to novel environmental conditions (Pfenning et al., 2016; Taylor et al., 2015). Therefore, it is imperative to employ eco-evolutionary models to comprehend the intricate interplay between environmental shifts and their genetic consequences in natural plant populations (Bush et al., 2016; Thuiller et al., 2013). For instance, Cotto et al. (2017) pioneered an innovative approach by integrating ecological niche modeling with individual-based genetic modeling to explore how plant populations in the Alpine range respond to evolving environmental conditions. Similarly, genetic offset methods were designed to predict the disparity between a population's current status and the necessary adaptations to be able to survive under future climatic scenarios (e.g., Rellstab et al., 2021). However, these approaches often fall short in fully accounting for the confounding effect of complex demographic history. For example, situations involving hybridization and strong stratification in population structures, such as well-defined populations belonging to different genetic clusters, are often not adequately addressed. Consequently, the full spectrum of both genetic and environmental variability exhibited by a species is frequently overlooked. This oversight is especially notable in continental-scale studies, where genetic diversity can also be influenced by gene flow across species. In our present study, we have taken a multifaceted approach by effectively merging ecological niche modeling, landscape genomics, and genotype–environment association analyses. Our aim is to unravel the pivotal role of recurrent hybridization in facilitating the adaptation of two cornerstone species within the boreal forest ecosystem to the challenges posed by global environmental changes.

In Western Europe, recurrent cycles of population contraction and expansion following glacial cycles shaped both genetic diversity (Milesi et al., 2023; Petit et al., 2003) and local adaptation patterns along environmental gradients that also correspond to post-glacial recolonization routes (Li et al., 2022; Savolainen et al., 2013; Thörn et al., 2021). For instance, in Norway spruce (Picea abies), height, diameter, and bud-set exhibit variations along latitudes, aligning with temperature and photoperiodic gradients (Chen et al., 2012, 2021; Milesi et al., 2019). Similarly, in P. obovata, phenology demonstrates latitude-dependent variations (Li et al., 2019). Conversely, bud-burst, a trait associated with phenology, shows changes along longitude, reflecting the influence of continental climate in P. abies (Milesi et al., 2019). Genetic admixture following range expansion from refugia can serve as a major source of genetic diversity in forest trees (Chen et al., 2019; Lexer et al., 2010; Müller et al., 2019; Tsuda et al., 2017; Zhou et al., 2023). It can promote increased climate resilience in wild populations (e.g., De Meester et al., 2018; Inoue & Berg, 2017) and aid wider phenotypic plasticity against rapid climate changes (see Catullo et al., 2019). Admixture caused by recurrent colonization by superior ecotypes can promote better resistance to environmental stress (e.g., Hamilton & Miller, 2016; Ortego et al., 2015) notably through an effective population size increase, persistence at the range edge and niche expansion (e.g., Gillies et al., 2016; Teukeng et al., 2022; Zalapa et al., 2010). Hybridization and admixture are thus effective mechanisms to increase overall genetic diversity and can further promote the transfer of adaptive alleles from one genetic background to another, enhancing local adaption and allowing populations to respond quickly to selective pressures (Rendón-Anaya et al., 2021; Rieseberg & Willis, 2007; Stelkens et al., 2014; Taylor et al., 2015). Here, we hypothesize that the recurrent hybridization and admixture events between two closely related coniferous species through cycles of historic range shifts due to ice cycles may help both species to better cope with future climatic changes.

Norway spruce [Picea abies (L.) H. Karst] and Siberian spruce (Picea obovata Ledeb.) are two closely related conifer species that diverged several million years ago (Chen et al., 2019; Tsuda et al., 2016; Zhou et al., 2023). Today, the two species have clear western (Western Europe) and eastern (Siberian Russia) distributions, albeit with a known hybrid zone located between the west of the Ural Mountains and expanding up to Northern Fennoscandia (Tollefsrud et al., 2015; Tsuda et al., 2016; Zhou et al., 2023; See map Figure 1). Currently, the distribution of P. abies is subdivided into seven main genetic clusters (Chen et al., 2019; Li et al., 2022). They correspond to three ancestral clusters (Alpine, Carpathian, Fennoscandian) as well as to admixture between these three main clusters (Chen et al., 2019; Li et al., 2022). Siberian spruce expanded westwards from Northeast Asia to Western Russia. Following glacial cycles, the two species came into contact on several occasions, leading to introgression (e.g., Chen et al., 2019; Tollefsrud et al., 2015; Tsuda et al., 2016; Zhou et al., 2023). Recently, Noto et al. (2022), using ancient DNA, and Li et al. (2022), using contemporary nuclear markers, confirmed conclusions drawn from pollen records (e.g., Giesecke, 2005) that recolonization of the Scandinavian peninsula by Norway spruce after the last glacial maximum (LGM) occurred through two different entry points, a southern one from Eastern location across the Baltic Sea and a northern one through northern Finland with a contribution from P. obovata. Li et al. (2022) demonstrated that the contact zone...
between the two main genetic clusters aligns with the boundary between Scandinavia’s primary climatic regions, revealing the influence of natural selection in maintaining this zone and shaping spruce genetic diversity. While pollen records are valuable for inferring species distribution and population movement during the Holocene (e.g., Trondman et al., 2015), they offer limited insights, particularly at the species level, for events predating that period, as observed in the case of Norway and Siberian spruce (Chen et al., 2019; Tsuda et al., 2016; Zhou et al., 2023).

We here used a combination of genome scans, genotype–environment association studies, and ecological niche modeling (for both past and future scenarios) to demonstrate that recurrent
hybridization significantly contributes to the expansion of both species’ niche breadth, potentially enabling them to better cope with changing climates. Given the close genetic affinity, the extensive hybrid zone, historic recolonization events, and the climatic gradient occupied by Norway and Siberian spruce, they provide a unique system for studying the role of hybridization in shaping climate resilience through genetic variation and range expansion. Our study first provides evidence of local adaptation and identifies key climatic variables influencing species distributions. We then use environmental niche models with these climatic variables and species occurrences to define the ecological niche preferences of both species and their historical movements. Finally, we integrate genetic and ecological information into genetic offset predictive models to assess the resilience of each population to future climatic conditions.

### 2 | MATERIALS AND METHODS

#### 2.1 | Sampling, genetic material, and SNP calling

The samples come from a large-scale study uncovering the demographic history and spatial population structure of the closely related Norway spruce (*Picea abies*) and Siberian spruce (*Picea obovata*) species (Zhou et al., 2023). DNA was extracted from the needles of 546 individual trees coming from 55 populations covering most of the distribution range of *P. abies* and the eastern range of *P. obovata* (including introgressed individuals and putative hybrid populations, following Chen et al., 2019 and Li et al., 2022: based on genetic data and Nakvasina et al., 2019: based on phenotypic data, see map Figure S1a). Note that the individuals from populations labelled R20, R22, and R24 in Figure S1a are also part of the provenance tests (equivalent to a common garden in breeding programs) analyzed by Nakvasina et al. (2019). Genomic DNA was extracted for NextGen sequencing using Qiagen® plant DNAeasy kit. Paired-end Illumina libraries were constructed and sequenced by RAPiD Genomic, United States, using 40,018 exome capture probes targeting 26,219 genes (Vidalis et al., 2018). Raw reads can be found on SRA (https://www.ncbi.nlm.nih.gov/sra) under PRJNA511374 and PRJNA1007582 meta information of populations (e.g., GPS coordinates).

Clean sequence reads were aligned to the *P. abies* reference genome (Nystedt et al., 2013) with BWA (Li & Durbin, 2009), and individual genotypes were identified using GATK HaplotypeCaller (Van der Auwera & O’Connor, 2020) followed by SNP calling across all samples with GenotypeGVCFs and hard-filtering of SNPs (QD < 2.0, MQ < 40.0, SOR > 3.0, QUAL < 20.0, MQRankSum < -12.5, ReadPosRankSum < -8.0, and FS > 60.0). For population genetic analyses, we retained only putatively neutral SNPs (i.e., synonymous mutations, mutations located in the introns or into the intergenic space) and pruned the dataset for singleton and SNPs in high linkage disequilibrium (plink v1.9, www.cog-genomics.org/plink/1.9/, pairwise $r^2 > 0.2$, window size 1kb), resulting in a 390,903 SNPs dataset. For all other analyses, we kept all SNPs with minor allele frequency higher than 0.05 including both neutral and non-neutral alleles (450,000 SNPs).

#### 2.2 | Climatic data selection

Environmental variables for current conditions were downloaded from the WorldClim 2.0 database at 30 arc seconds resolution (Fick & Hijmans, 2017; Hijmans et al., 2005). They are annual averages of both temperature and precipitation-related variables. Climatic records at population locations were extracted with the TERRA R package (Hijmans et al., 2022). In order to avoid multicollinearity and model overfitting, we performed pairwise Pearson correlation between all the bioclimatic variables, sequentially removing the variable with the highest correlation coefficient, $r$, until it reached $r_{\text{max}} = 0.6$. Additionally, to ensure that multicollinearity is minimal in the composite variables of RDA, we assessed the variance inflation factor (VIF) and retained the environmental variables for which the VIF was lower than five, as recommended by Capblancq and Forester (2021). We used this final set of environmental variables (predictor variables hereafter) for all downstream analyses.

#### 2.3 | Hybrid index

We calculated a hybrid index using the ADMIXTURE software (v1.3.0; Alexander et al., 2009) assuming $K=2$ ancestral populations. The hybrid index was defined as being the proportion of *P. abies* ancestry component and ranges from 0 “pure” *P. obovata* to 1 “pure” *P. abies*. The population-wise hybrid index was obtained by averaging individual hybrid index. Populations with a hybrid index higher than 0.2 and below 0.8 were classified as hybrid populations in qualitative analysis. Extrapolation of the hybrid index to the whole distribution of the two species was done using custom-R scripts (available in the GitHub repository) with nearest-neighbor interpolation where the imputation essentially calculates the estimated hybrid index at a specific location by taking the mean of the known hybrid index observed in the nearby locations within a defined neighborhood. The interpolation was mapped to visualize the spatial distribution of hybrid index of the two species.

We then investigated the relationship between the hybrid index and various metrics of genetic diversity and variation in allele frequencies. In order to minimize the skewed allele frequencies due to the varying number of samples in each population, we (i) randomly sampled 10 individuals per population for populations with more than 10 individuals and (ii) grouped close-by populations that were also genetically similar before downsampling to 10 individuals. The resulting dataset consisted of 36 populations.

We calculated global nucleotide diversity ($\pi$: Nei & Li, 1979) to assess population-level genetic diversity with the POPGENOME R package (Pfeifer et al., 2014). Extrapolation of the $\pi$ values to the
whole distribution of the two species was done using the same approach as for the hybrid index.

2.4 | Genetically informed climatic group definition

In order to find genetically informed “climatic groups” of both species, we combined two sets of data: (i) adaptively enriched genetic space scores from pRDA—the RDA scores represent local adaptation in populations derived from allele frequencies, (ii) bioclimatic data—this dataset represents the environmental variability of populations after removing of multicollinearity. The former was used as a proxy for population-level adaptations to local climatic conditions and the latter to explain the spatial variation of local conditions. A k-means clustering was performed for the total within the sum of squares to determine the optimal number of genetically informed climatic clines. This was performed separately for tentative hybrid and non-hybrid populations to avoid cluster overlaps that could arise from specific adaptations in hybrid populations. In non-hybrid populations, we identified five distinct climatic groups while there were only three within putatively hybrid populations. After obtaining the optimal number of clusters as five, a second k-means clustering was performed with \( K = 5 \) for all populations (i.e., including hybrid and non-hybrid populations) to determine the climatic group membership of each population. This step was performed specifically to group the putative hybrid populations with their closest non-hybrid populations to identify their closest climatic groups.

2.5 | Identification of alleles putatively involved in local adaptation

To study the genetic architecture of local adaptation in Norway spruce and Siberian spruce, we ran genome scans to detect outlier loci (i.e., loci showing an excess of genetic differentiation in comparison to the rest of the genetic background) using the principal component-based outlier detection method, “pcadapt,” implemented in the R package “pcadapt” v.4.3.2 (Luu et al., 2017; Privé et al., 2020). Briefly, a Z-score matrix, obtained from regressing all SNPs to PCs, is used to produce one multivariate distance for each variant. Assuming that SNP variation along principal component axes largely reflects demographic processes, this method allows controlling for population structure when testing for local adaptation. In our analysis, to control for false-positive risk discovery associated with multiple testing, we used FDR \( q \)-value < 0.1 to obtain the significant SNPs putatively involved in local adaptation.

Redundancy analysis (RDA) is an asymmetric canonical analysis that combines ordination and regression, where multiple regression on the response variable (i.e., allele frequencies) and explanatory variables (climatic variables) is followed by ordination of fitted values of the regression (van den Wollenberg, 1977). We used redundancy analysis (RDA) to find the main climatic variables that explain most of the variance in allele frequencies between the two species and to identify candidate loci associated with environmental variation.

In the special case of partial RDA (pRDA), additional explanatory variables (covariates) are used to control for various confounding effects (see Capblancq et al., 2018; Capblancq & Forester, 2021). This aspect of RDA is extremely useful in association analyses to control for the confounding effect of factors that are not the focus of the study, such as population structure and geographic distance. Thus, to isolate different drivers of genetic variation, in addition to a full model (including all predictor variables), we performed three independent pRDA models controlling for geographic distance, population structure, and environmental conditions. We adopted the method described in Capblancq and Forester (2021) where the Mahalanobis distance of each locus to the center of the RDA is calculated using two axes (\( K = 2 \)) of partial RDA (i.e., the number of axes that explain more than 60% variation; following Capblancq & Forester, 2021), followed by obtaining the top outlier SNPs from this distance matrix. For geographic distance, the cartesian distance of populations was used as a proxy; to control for population structure, we used the first three axes of a PCA performed on the set of putatively neutral loci.

Using the output from the pRDA, we computed an Adaptive index \( A_{i,j} \) as implemented by Capblancq and Forester (2021) from Steane et al. (2014) using the RDA 1 and 2 independently. For the top outliers, RDA1 and RDA2 explained 88% and 10% of the variation, respectively. For each location, \( j \), the adaptive index is given by the sum over each climatic variable, \( i \), of the products between its score along a given RDA axis and its value at location \( j \).

\[
A_{i,j} = \sum_{i=1}^{n} a_{i} b_{j}
\]

where \( a_{i} \) is the climatic variable score along the RDA axis, \( b \) is the standardized value for that particular variable at the focal site (raster cell). \( A_{i,j} \) ranges from negative to positive values and essentially measures the match between the genotypes and their environment, 0 meaning no match.

2.6 | Niche expansion, species distribution modeling, and historical niche optima movement analysis

The climatic niche of the two species was defined as the ordination space of all the populations with PCA eigenvalues. Therein, a PCA was performed on the selected climatic variables (i.e., Bio8, Bio9, Bio18) for all the populations of \( P. abies \) and \( P. obovata \) separately. The optimal effect of distance was determined as the centroid of the eigenvalues and the niche breadth was defined as the 95% ellipsoid of the PCA scorings. The analysis was performed with and without hybrid populations to visualize the niche breadth overlap. As a complementary analysis, we also used a non-parametric method, collective environmental gradient (CEG; Karunarathne et al., 2018) to assess the total environmental aptitude of each species and to visualize their
overlap. CEG is a single-value decomposition approach to determine the composite two-dimensional environmental/niche breach. We performed CEG according to Karunarathne et al. (2018) for the selected bioclimatic variables (i.e., Bio8, Bio9, Bio18).

Species distribution models for the current, mid-Holocene (ca. 6000 ya), last glacial maximum (LGM, ca. 22,000 ya), last interglacial period (ca. 120,000−140,000 ya), and future climatic scenarios (2021−2080 and 2081−2100) were constructed using maximum entropy (Maxent) modeling (Phillips et al., 2006). The optimal niche for separate species was determined with selected climatic variables (i.e., Bio8, Bio9, Bio18) using the maxent function of the dismo R package (Hijmans et al., 2016). Maxent uses presence-only data to estimate the probability distribution of occurrence based on environmental constraints (Phillips et al., 2006). In order to test the models and to facilitate a more accurate prediction of species distribution, we downloaded species occurrence data for P. abies and P. obovata from GBIF (www.GBIF.org) and filtered the data to include only confirmed and natural occurrences of the species. Occurrence data were then subsampled to retain only one occurrence per grid cell of 1 arc minute. The final dataset consisted of 619 occurrences (P. abies: 288; P. obovata: 331); hybrid populations have been excluded. All bioclimatic variables were downloaded from WorlClim 2.0 database at 30 arc-second resolution. The climatic variables for the Mid-Holocene, LGM, and last interglacial were downloaded from the model CCSM4 and the future climate variables were downloaded for ACCESS-CM2 model. The projection of past and future niche distribution of the two species was done using the predict function of the dismo R package. All the models were evaluated using the area under the operating curve (AUC) and only the models with an AUC of at least 0.85 were considered.

2.7 | Assessment of genetic offset and resilience to future climatic conditions

The genetic offset of an entity (e.g., population) is the allele frequency difference between the current and required frequency in the future to survive in future climatic scenarios. We calculated the genetic offset of our populations with risk of nonadaptedness (RONA) as in Hellstab et al. (2016). RONA is the average theoretical change in allele frequency at climate-associated loci required to match future climatic scenarios. This is established by regressing the selected loci allele frequencies with the selected environmental variables and using the regression coefficients to predict the expected allele frequencies of the respective loci to future conditions. The difference between the two allele frequencies is the RONA. Here, we first performed linear regression on all loci with the selected environmental variables (i.e., Bio8, Bio9, Bio18) and kept only the loci that showed significant association (p-value <0.05) with the three environmental variables. Then, we obtained the intersection between these alleles and the top outliers from the RDA analysis for the RONA analysis. To avoid linkage between selected loci, we filtered loci within 10,000 base pairs in the same scaffold; the locus with the highest $R^2$ from the regression was retained. In total, ~25,000 candidate SNPs for environmental associations were kept for the analysis. Finally, we used nearest-neighbor interpolation (as in Rukundo & Cao, 2012: custom scripts available in the GitHub repository) to predict the RONA values for the entire distribution of the two species for spatial visualization.

3 | RESULTS

3.1 | Populations exhibiting a high level of admixture are spread across three climatic groups

We used the same sampling of 55 populations covering the whole range of P. abies and the western part of the distribution range of P. obovata as in Zhou et al. (2023). Using a combination of population genetics analyses, the authors showed that there are two main genetic entities: one western cluster, corresponding to Norway spruce (P. abies), centered on the Alps and the Carpathians, and extending toward Scandinavia and the Russian Plains, and one eastern cluster stemming from Siberia, east of the Urals Mountains and corresponding to what has generally been recognized as Siberian spruce (P. obovata). The populations previously recognized as of hybrid nature located west of the Ural Mountains and east of the Baltic region formed subclusters within the two main clusters (Figure S1a, also see Zhou et al., 2023: Figure S3). Following this initial clustering, in order to quantify the degree of admixture between P. abies and P. obovata, we derived a hybrid index for each individual using ADMIXTURE $(K=2$, see Section 2 and Figure 1b; Figure S1b) and averaged it per population. The distribution of the hybrid index is highly consistent with the Bayesian clustering in Zhou et al. (2023). To investigate whether specific DNA polymorphisms were associated with given climatic zones, we used redundancy analysis (RDA) incorporating both genetic and climatic data (see Genetic Diversity Distribution Among Populations in Section 2). We defined five major environmental clusters (i.e., genetically informed climatic groups) among all the non-hybrid populations while populations in the hybrid zone split into three different clusters (Figure 1a). Interestingly, this divided western populations (i.e., populations west of the hybrid zone, recognized as P. abies) into three groups: a southern group consisting of Alpine populations, a northern group made of Fennoscandian and Baltic populations, and a third group with Carpathian and Central European populations, while eastern populations (i.e., populations east of the hybrid zone, recognized as P. obovata) were separated into a northern and a southern group along the southern border of the west Siberian taiga (Figure 1a).

3.2 | Spatial heterogeneity and environmental conditioning of genetic diversity

We investigated changes in total nucleotide diversity ($\pi$) across the whole distribution range of the two species. Globally and in line
with recent estimates (Chen et al., 2019; Milesi et al., 2023; Zhou et al., 2023), x ranged from 0.00518 to 0.00695, the highest values were found in the north-eastern range for *P. obovata* and Southern Scandinavia for *P. abies*. The populations located in the contact zones (e.g., in Scandinavia between the Northern and the Southern range) or located within the large hybrid zone between *P. abies* and *P. obovata* exhibit higher x values than neighboring populations, exhibiting the effect of introgression and/or admixture on genetic diversity.

As expected, hybrid populations also show intermediate heterozygosity between *P. abies* and *P. obovata* (linear model, F = 6.55, df = 34, p = .015, Figure 2b). More importantly, hybrid populations tend to have more homogeneous allele frequencies from one SNP to another than in *P. abies* (Welch t-test, t = −2.85, df = 11.172, p = .02) or *P. obovata* (t = −5.63, df = 7.93, p < .001, Figure 2c): Note that standard deviation in allele frequencies varies quantitatively with the hybrid index, meaning that both standard deviation and heterozygosity are at a minimum for a hybrid index closed to 0.5 (Figure S4a,b).

### 3.3 Local adaptation

We further investigated whether the observed pattern of genetic differentiation could be due to local adaptation using *pcAdapt*, a genome scan method to detect outlier SNPs exhibiting a higher degree of differentiation than the genetic background while controlling for population structure. We detected 1650 SNPs, clustered into specific areas of the genome (Figure S3a,c) suggesting that natural selection has shaped the current genetic diversity, besides the population movements and demographic history.

In order to understand the role of environmental variation in shaping genetic diversity, we conducted a redundancy analysis (RDA) with bioclimatic factors as predictors (Figure S3b). After an ad hoc selection to remove the high levels of correlation between climatic variables, we retained three climatic variables: Bio8, the mean temperature of the wettest quarter; Bio9, the temperature of the driest quarter; and Bio18, the precipitation of the warmest quarter. While Bio8 mainly varies along latitudes, Bio9 and Bio18 mainly reflect continentality and vary longitudinally. Together the climatic variables explained 5.4% of the total genetic variation in the climatic partial RDA (pRDA) where co-variables were introduced to control for population structure (Figure S3b). Still, we detected 1536 outlier loci significantly associated with at least one of the bioclimatic variables; out of which 298 were recognized as top outliers, based on the overlap of a simple RDA model and the climatic model (see Capblancq & Forester, 2021 for details). These top outliers clustered together in various regions of the genome (Figure S3a). Outlier SNPs showed substantial congruence with the enrichment observed in the *pcAdapt* output (82% overlap, Figure S3a,d) suggesting that a significant part of the allele frequency variation observed among populations is driven by environmental heterogeneity. From the pRDA, we derived an *adaptive index* (see Materials and methods) that measures the match between the genotypes and their environment to capture the spatial distribution of population-level adaptation to current climatic conditions (Figure 3). The adaptive index and the hybrid index were highly correlated (Spearman’s rho = 0.90, S = 2156, p < .001 and rho = 0.40, S = 13,339, p = .003, for adaptive index computed along RDA1 and 2, respectively). While the populations at both extremes of the sampling range show the highest degree of adaptation (strongly positive or negative values for the adaptive index), populations located within the hybrid zone exhibit intermediate values and close to zero indicating a lower degree of local adaptation to these climatic variables (Figure 3a–c).

We subsequently tested whether the observed pattern is primarily influenced by candidate SNPs specific to the environmental variables included in the pRDA or if it demonstrates a broader, overarching trend. For each population, we calculated the proportion of

![Figure 2](https://example.com/figure2.png)

**Figure 2** (a) The distribution of nucleotide diversity (π) throughout the complete distribution range of the two species (points show the non-extrapolated values). (b, c) Depict population-wise heterozygosity and average and standard deviation in allele frequencies across SNPs for *P. obovata*, *P. abies*, or hybrid populations. Map lines delineate study areas and do not necessarily depict accepted national boundaries.
major alleles specific to that population. For all SNPs individually, we initially identified the minor allele as the one with the lowest frequency across the entire dataset ($f_{\text{all}} < 0.5$). We then determined, for each population individually, the proportion of SNPs for which the minor allele was different than across the entire dataset. In other words, while an allele has the lowest frequency globally, its frequency within the focal population is the highest (i.e., $f_{\text{all}} < 0.5$ but $f_{\text{pop}} > 0.5$). First, the hybrid populations showed a significantly lower proportion of population-specific major alleles than $P. abies$ or $P. obovata$ populations (linear model, $r^2 = 0.65$, $df = 33$, $p < .001$, Figure 3d; Figure S4a). Second, a strong positive correlation exists between the absolute value of the adaptive index of a population and its proportion of population-specific major alleles (Spearman’s $\rho = 0.67$, $S = 2532$, $p < .001$, also see Figure 3e); the same relationship, albeit weaker, is observed when computing the adaptive index along RDA2 (Spearman’s $\rho = 0.33$, $S = 2502$, $p = .05$). The lower adaptive index in hybrid populations is not solely attributed to markers associated with the environmental variables considered in the pRDA analysis. Instead, it is indicative of widespread shifts in allele frequencies throughout their genome (Figure S4a,b). Additionally, $P. obovata$ populations, characterized by a greater number of population-specific major alleles, tend to exhibit higher overall genetic diversity.

### 3.4 Environmental preference, historical niche movement, and genetic refugia

We then used an ordination of climatic ranges of each species from the occurrence data to define the niche envelopes of the two species and contextualize their niche optima. The range of environmental preferences varied substantially between the two species albeit with visible overlap (CEG: non-parametric environmental gradient Figure 4a). Interestingly, the peaks of the two niches were far from each other in our environmental gradient analysis although the complete range of the niches showed considerable overlap. The PCAs of climatic niche space also revealed an overlap between the two species, suggesting that marginal populations of both species close to the hybrid zone are adapted to similar environmental conditions (Figure 4b). Further, when
hybrid populations are considered in the analysis, the overlap becomes wider, suggesting niche expansion owing to hybridization. The PCA ellipsoid area increase (1.5 times for *P. abies* and 2.3 times for *P. obovata*) in Figure 4c shows this expansion.

To shed light on spatial variation in nucleotide diversity, we used niche modeling to infer species’ suitable habitat during LGM (~25,000 years ago) and mid-Holocene (~4000–10,000 years ago Figure 5). It confirmed that the southern range of *P. abies* (e.g., Alpine, Carpathian, and western European forests) harbored suitable habitat and must have served as refugia during the last glacial maximum. With the warming climate, the species migrated back to higher latitudes, eventually establishing its present distribution. Niche modeling for the period before the LGM, and specifically for the interglacial period (i.e., ~140,000 years ago), shows that optimal habitats for *P. abies* existed in the northernmost regions of the current distribution (e.g., northern Sweden and Norway) as well as in the Alpine region. Probing that the distribution of Norway spruce followed the movement of its niche during cycles of glaciation, population connectivity of *P. abies* must have been drastically affected. In contrast, *P. obovata* may have maintained a broader distribution compared to *P. abies*, in particular during the LGM (see the discussion of “Recurrent hybridization” for genetic support). Notably, the suitable habitat ranges of both species frequently intersected over the course of the glacial cycle, with a notable convergence observed in the western Russian forests during LGM. This overlap could have facilitated multiple instances of hybridization and genetic exchange between the two species. This phenomenon mirrors the current distribution scenario where a modest spatial overlap of the suitable ecological niche is evident in North European Russia (sensu TDWG classification) (Figure 5).

3.5 Genetic offset and resilience to climate change

Lastly, we explored the capacity of present-day populations of both species to respond to climate change. To achieve this, we quantified population adaptability to future climates by assessing the discrepancy in allele frequencies between current and projected values based on anticipated climatic conditions, referred to as the “genetic offset.” RONA (Rellstab et al., 2016) measures the allele frequency change necessary for future climate based on a linear regression of current values. We calculated RONA for three bioclimatic variables using climate projections for the years 2080 and 2100 (i.e., Bio8, Bio9, and Bio18). The linear responses of the frequencies of the top 10 alleles of RONA were remarkably significant (p-value <.001) for the respective environmental variable (Figure S5), indicating high predictability of adaptedness, especially the lack thereof. Overall, the RONA assessment indicates a relatively low RONA to all three
Figure 5 Distribution of suitable habitat/niche for *P. abies* and *P. obovata* from the last interglacial period to future climatic conditions, showing the niche movement from the past to present, determined by maxent modeling of the niche optima. Map lines delineate study areas and do not necessarily depict accepted national boundaries.
bioclimatic variables in regions characterized by high nucleotide diversity (Figure 6). This underscores that nucleotide diversity and likely overall genetic diversity serves as a pivotal factor in mitigating the risk of non-adaptedness, thereby providing populations with a sheltered source of adaptive potential. We observed high RONA scores in populations from central Europe and at the Southernmost range of *P. abies* (0.18–0.22) indicating that populations in these regions may be at higher risk of extinction under future climatic conditions. However, this trend was not seen in populations in the Alpine region which has served as a refugium during the glacial period and also most likely in warmer conditions during interglacial periods (see Figure 5).

Even more intriguingly, in the case of both species, the RONA significantly reaches its nadir within the hybrid zones (Spearman’s correlation between average RONA values and folded hybrid index, \( r = -0.67, S = 46,405, p < .001; \) see also categorical analyses, Figure 6, bottom right panels). This observation further bolsters the proposition that admixture amplifies genetic diversity, thereby fostering the broadening of ecological niche range, a phenomenon conducive to species adaptation to future climatic scenarios. Despite the fact that the Siberian spruce populations harbor the broadest genetic diversity overall (with a \( \pi \) range of 0.0059–0.0069), they paradoxically exhibit a greater susceptibility to non-adaptedness compared to the hybrid populations. Finally, we integrated insights from the genotype–environment association study with those of the RONA analysis; about two-thirds (\( N = 194 \)) of the SNPs detected as outliers in the pRDA are among the SNPs with the highest RONA values. This convergence signifies that the SNPs responsible for local adaptation to current climatic conditions are potentially detrimental in the context of long-term response to climate change as highlighted by the positive correlations between the absolute adaptive index and the average RONA (Spearman’s \( r = 0.43, S = 15,647, p = .001, \) Figure S6; Table S1); the higher the degree of adaptation to the current environment, the higher the RONA to future environments.

4 | DISCUSSION

Extensive evidence now highlights the detrimental influence of global climate change on biodiversity at multiple scales (see Altieri &
Gedan, 2015; Walther et al., 2002; Williams et al., 2021). Population-level genetic variation equips species with the capacity to navigate shifting climates, safeguarding them from extinction (De Meester et al., 2018). This is often observed through species’ adaptations to their local environment and reflects their potential for future climate change resilience (Sang et al., 2022). Nevertheless, identifying these environment-linked alleles necessitates extensive whole-genome sequencing of multiple populations and/or long-running common garden experiments. Nonetheless, this approach proves impractical for species characterized by extended generation times, frequently overlooking essential evolutionary parameters such as demographic history and the extent of gene flow. In the present study, using exome capture sequences, we investigate the role of species hybridization in shaping the current ecological niches of Norway and Siberian spruce. Our findings reveal that recurrent admixture has heightened the climate change resilience and adaptive capacity of both species. Additionally, this approach has enabled us to pinpoint populations vulnerable to prevailing climate change conditions.

4.1 | Niche modeling suggests recurrent shifts from southern to northern distribution ranges

Our analysis of niche optima movement of Norway spruce during and post-LGM agrees with fossil records and ancient DNA analyses (Brewer et al., 2017; Nota et al., 2022; Parducci et al., 2012; Tollefsrud et al., 2015). Tollefsrud et al. (2015) traced the optimal ecological niches of Norway spruce and Siberian spruce during the most recent glaciation cycle. They also delineated the distribution ranges of these species during the LGM and the mid-Holocene. The study’s findings showcased the gradual migration of both species to and from their southern refugia. This is of course expected since many species in temperate regions followed the same path during to and from their southern refugia. This is of course expected since many species in temperate regions followed the same path during the last glaciation cycle (Feliner, 2011). Our analysis of niche optima movement of Norway spruce during and post-LGM agrees with fossil records and ancient DNA analyses (Brewer et al., 2017; Nota et al., 2022; Parducci et al., 2012; Tollefsrud et al., 2015). Tollefsrud et al. (2015) traced the optimal ecological niches of Norway spruce and Siberian spruce during the most recent glaciation cycle. They also delineated the distribution ranges of these species during the LGM and the mid-Holocene. The study’s findings showcased the gradual migration of both species to and from their southern refugia. This is of course expected since many species in temperate regions followed the same path during the last glaciation cycle (Feliner, 2011). However, inferring P. abies and P. obovata ecological niches further back in time (~140 k years ago), a period where macrofossils and/or palynological evidence fail to provide species-specific information, unveiled a substantial overlap in the distribution ranges of both species in their northern habitats. This convergence could have paved the way for multiple instances of admixture. This is highly consistent with Chen et al. (2019) and Zhou et al. (2023) estimates of a massive admixture event (~103–187 kya between P. abies and P. obovata populations using genome-wide SNPs. Our niche modeling approach further reveals that the Southern European Mountain ranges, including the Alps and the Carpathians, likely acted as refugia for P. abies during both glacial and interglacial periods. Meanwhile, the species’ primary distribution underwent a transition from southern to northern latitudes in tandem with the shifts in glaciation cycles. The recurrent utilization of the same discrete and well-defined refugia across ice cycles may, at least partially, explain the difference observed between Norway spruce and two other boreal tree species—Pinus sylvestris (Scots pine) and Betula pendula (Silver birch)—both of which display notably less population structure (Bruxaux et al., 2023; Milesi et al., 2023; Tsuda et al., 2017). The latter may have instead survived the glacial periods in a large metapopulation south of the glaciated regions (Hewitt, 2000; Palmé et al., 2003).

4.2 | Genetic diversity does not reflect niche centrality but captures distribution range variations

Overall, P. obovata populations exhibit higher \( \kappa \) values than P. abies, likely reflecting more stable historical distribution and more continuous distribution through space (Figures 2a and 5). In P. obovata, a clear shift in \( \kappa \) matches the location of the Ural Mountain, which likely limits the gene flow. Pockets of high diversity are found in the Southernmost range of P. obovata, west of the Urals, possibly reflecting cryptic refugia during the ice age and as suggested by Zhou et al. (2023). While the overall genetic diversity within P. abies also ranges between 0.0054 and 0.0068, the northern- and southern-most populations had the lowest nucleotide diversity. Even though southern locations in Italy and the Alpine mountains region have been refugia for Norway spruce during the ice ages, the current relative isolation of the populations, because of the topography, probably explains the relatively lower genetic diversity. The northernmost populations of P. abies in Sweden harbor the lowest genetic diversity. These populations are also the youngest and the low nucleotide diversity may reflect the colonization forefront (Li et al., 2022; Nota et al., 2022). Nevertheless, the highest \( \kappa \) values are found in Southern Scandinavia, an area that was also recolonized after the LGM but from different populations (Li et al., 2022; Nota et al., 2022).

Niche centrality suggests that a species’ performance is best at its optimal niche (also niche centroid) conditions (see Lira-Noriega & Manthey, 2014). Although the niche center of a species may or may not be the geographical center of the species, it hypothesizes that the genetic diversity in the niche center is higher than in the peripheral populations (Manel & Holderegger, 2013; Pironon et al., 2015). Our results, however, show little congruence with niche centrality expectations (Figure 5). For P. abies, we recognized southern Scandinavia as the niche centroid (Figure 4; see the populations in the center of eigen ellipses) and populations in this region show the highest nucleotide diversity (Figure 2a). It is, however, imperative to mention that the currently observed high genetic diversity of P. abies in the southern Scandinavian forests is also a result of recent introductions of genetic material from the whole range of the species (Chen et al., 2019 and reference therein). For P. obovata, our sampling does not represent the complete current distribution of the species, especially east of the Yenisei River. For the western range of the species, the highest genetic diversity is found along the Yenisei River and the niche centroid points toward the northern Ural Mountains. While acquiring a more continuous sampling would enhance the informativeness of our conclusions, various other factors could also account for the absence of support for niche centrality. First, both

---

For the complete reference list, please see the original publication. This summary is intended to provide an overview of the key points discussed in the text. For more detailed information, please refer to the original research article.
Norway and Siberian spruces, as boreal species with intricate demographic histories, experienced shifts in their suitable habitats and ecological niches across multiple ice-age cycles. For *P. abies*, populations were even extirpated during the LGM from what represents today the niche centroid. Additionally, the long generation time and long-range pollen dispersal of coniferous species could also contribute to the lack of fit to the Niche centrality expectations in these species, as is observed in *Pinus sylvestris* across its whole distribution range (Bruxaux et al., 2023). Our findings align with recent research that observed minimal variation in estimates of genetic diversity among populations from various dominant forest tree species in Western Europe (James et al., 2023; Milesi et al., 2023). Finally, we must acknowledge that although we use nucleotide diversity as a proxy for overall genetic diversity, it may not fully capture other aspects of genetic diversity, such as variation in the number of alleles, or structural variations in the genome.

### 4.3 Recurrent hybridization enlarged the ecological niche breadth

Our niche modeling strategy and retrospective projections offer additional ecological validation for the influence of hybridization on shaping the genetic diversity of both species. Our results are in agreement with genetic data, demographic modeling (Zhou et al., 2023), and paleoecological records (Semerikov et al., 2013 and references therein), which suggest Siberia's state during the last glacial cycle, and possibly more ancient glacial cycles, as a dry desert interspersed with pockets of forested areas. In agreement with these data, our niche modeling results (Figure 5) indicate that the suitable habitat of *P. obovata* remained notably larger than that of *P. abies* during the late Pleistocene and Holocene periods. Furthermore, it can be seen in Figure 5 that the suitable habitat for both species overlapped in many instances during the Pleistocene and Holocene, in particular in their northern range, representing opportunities for genetic introgression. Contrary to limited gene flow, Zhou et al. (2023) demonstrated that secondary contacts among distinct genetic entities of the two species amplified local genetic diversity. This phenomenon is evident in populations situated within present contact zones or within the expansive hybrid zone between *P. abies* and *P. obovata*, which displays high *π* values compared to neighboring populations (Figure 2). More importantly, Zhou et al. (2023) highlighted that the interactions between *P. abies* and *P. obovata* led to the emergence of novel genetic clusters with admixed characteristics. These clusters have endured through to the present day, persisting within cryptic refugia during glacial periods. Furthermore, they proposed that during the expansion phases, these entities helped to bridge *P. abies* and *P. obovata* genetic backgrounds, serving as stepping stones (see also Tsuda et al., 2016).

The role of hybridization in allowing niche expansion has been well documented in many cases of invasive species (e.g., Pfennig et al., 2016 and reference therein), particularly in the presence of long-distance gene flow (e.g., Berthouly-Salazar et al., 2013) that allows for the maintenance of genetic diversity in peripheral populations and for the introgression of favorable alleles from other populations. Hybridization can also directly contribute to favorable traits (e.g., Darwin’s finches’ beak shape, Grant & Grant, 2019) and foster their evolution through an increase in standing genetic variation, while adaptation from de novo mutations can be slow (Rieseberg, 1997; Taylor et al., 2015); especially in species with long generation time like spruces.

In Norway and Siberian spruce, primary phenotypic traits related to growth phenology or reproduction exhibit a continuous spectrum of variation between the two species. Hybrid individuals, in this context, manifest intermediate phenotypes (Lagerrantz & Ryman, 1990; Orlova et al., 2020; Popov, 2010). Categorized by their phenotypes, the hybrid populations can be distinguished into two forms, each exhibiting “abies” or “obovata” characteristics (Nakvasina et al., 2019). These distributions align with the delineation of the hybrid zone as outlined by Zhou et al. (2023) and precisely coincide with the distribution of the hybrid index generated in our study from genomic data (see Figure 2). Provenance tests (i.e., breeding experiments equivalent to common garden experiments) also revealed that the individuals of hybrid nature, characterized as such based on phenotypic characteristics, also had higher survival rates in the hybrid zone (Nakvasina et al., 2019). Our study confirmed with genetic data their hybrid nature (see also Zhou et al., 2023). Consistent with these findings, our study also demonstrates that the hybrid populations do not necessarily fall within the ecological niches of either species. Instead, they carve out a niche of their own that partially intersects with the ecological preferences of both species (see Figure 4). Therefore, it is likely that hybridization is responsible for expanding the ecological niche breadth of both species.

### 4.4 Integrating niche modeling, population genetics, and genetic-offset analysis

Given the current pace of global climate change, the central question is, will long-lived organisms such as spruces have enough time to adapt to the new environmental conditions before facing an inevitable demographic collapse? First and foremost, our niche modeling to future scenarios suggests that suitable habitats would be available for both species in their current northern distribution range and our genetic offset analysis indicates that these populations should be poised to counter climate change (as indicated by a low-risk of non-adaptedness, see Figures 5 and 6). However, it is worth noting that a considerable portion of the current range of *P. abies*, as well as a substantial segment of the western distribution range of *P. obovata*, might not be suitable for their continued survival.

The populations situated at the opposite ends of the adaptive index, as indicated by the genetic offset analysis, correspondingly represent those populations most susceptible to future climate
conditions (Figures 3a and 6). This observation aligns with the expectation that populations displaying a pronounced degree of adaptedness can be regarded as “specialists” tailored to their specific environments. However, according to our modeling of the future distribution ranges, regions with populations showing an elevated RONA are projected to accommodate suitable habitats for species persistence (Figure 5). At first glance, these two results may seem contradictory. The essence of the RONA analysis lies in gauging the average change in allele frequency of SNPs associated with key climatic variables. This change is required to maintain an equivalent level of adaptedness under future climatic scenarios. This suggests that the alleles that are responsible for local adaptation to the current climate might potentially be deleterious under future environmental conditions, and consequently, they could be at risk of being lost. However, RONA, like many other genetic offset prediction models, suffers from well-identified limitations, including its failure to account for demography history, population structure, or gene flow (Gain et al., 2023; Rellstab et al., 2021).

Both P. abies and P. obovata exhibit extensive census sizes and effective population sizes alongside far-reaching pollen dispersal (Zhou et al., 2023). This characteristic could potentially aid in the dissemination of adaptive alleles, particularly considering the distinct ecological preferences of the two species (see Figure 4). Our study has further revealed that hybrid populations occupy their own ecological niche while having a unique genetic makeup (i.e., a lower fraction of population-specific major alleles and more even frequencies from one SNP to another). Moreover, these hybrid populations could serve as potential sources for future recolonizations, in addition to acting as a reservoir of genetic diversity. Both species hold significant economic importance, and human-assisted transfer of trees from different provenances could play a role in bolstering adaptation to future conditions. For instance, P. abies trees of Southern origin planted in Southern Scandinavia for breeding purposes (Chen et al., 2019) exhibited higher growth rates in Southern Sweden compared to trees of local origin (Milesi et al., 2019). Intriguingly, Southern Sweden presently showcases the highest genetic diversity in P. abies (Figure 2a). Consequently, these trees can potentially intragress with local counterparts leading to the introduction of favorable alleles for future climatic conditions. Notably, the plants’ response to global change can be rapid, even in long-lived species, like oak (Quercus petraea), which despite its ~50-year generation time, displayed a genome-wide response to the Little Ice Age (~1650 AD, Saleh et al., 2022).

Finally, while neither of the studied spruce species might face a large risk of global extinction under future climatic conditions, it is worth noting that certain areas are under particular threat. Among populations with the highest risk of non-adaptedness, two are located south of the Italian Alps and one, Indigo, in the Northern range of P. obovata western domain. As highlighted by Zhou et al. (2023), these populations are geographically isolated from the rest, exhibit low nucleotide diversity, and have undergone more pronounced genetic drift. These characteristics often point to populations that might struggle to contend with the challenges posed by global climate change.

4.5 Conclusion and perspective

When species are unable to track their ecological niche, standing genetic variation is the main determinant for species’ ability to cope with climate change. This is even more true for species with long-generation time where the probability of adaptation through de novo mutation is low. Cycles of range contractions and expansions can enhance genetic diversity in a feedback loop (Barrett & Schluter, 2008; Sexton et al., 2009) albeit with expected diversity loss during contraction events (Pauls et al., 2013). In Western Europe, species-wise genetic diversity of seven forest tree species globally increased across multiple ice cycles (Milesi et al., 2023).

Among them, at least five (Betula pendula, Quercus petraea, Fagus sylvatica, Populus nigra, and P. abies) are known to hybridize with closely related species and are all long-lived species with long-range gene flow. Hybridization could thus be a main factor in coping with climate change in temperate and boreal forest tree species, as our study suggests in the P. abies/P. obovata species complex. There is a burst of new methodology development aiming at integrating niche modeling and genetic offset prediction in order to better apprehend species response to global change (e.g., Barratt et al., 2023; Chen et al., 2022). If this clearly represents a step forward, our study also evidenced the need for a more integrative framework to also consider species demography history, population structure, and gene flow.

AUTHOR CONTRIBUTIONS

Piyal Karunarathne: Conceptualization; data curation; formal analysis; methodology; writing – original draft; writing – review and editing. Qiujie Zhou: Data curation; formal analysis; writing – review and editing. Martin Lascoux: Resources; validation; writing – review and editing. Pascal Milesi: Conceptualization; data curation; funding acquisition; methodology; resources; supervision; writing – review and editing.

ACKNOWLEDGMENTS

We would like to thank Elena Nakvasina, Vladimir Semerikov, Lars Opgenoorth, Katrin Heer, and Giovanni Giuseppe Vendramin for their valuable efforts in providing us with plant materials. We also thank the Swedish National Infrastructure for Computing (SNIC) for resource allocation for high-power computing under the project numbers SNIC 2021/6-313 and SNIC 2021/5-540. The Nilsson-Ehle grant by the Royal Physiographic Society of Lund awarded to Piyal Karunarathne (42191) was immensely helpful for acquiring genome sequences to fill data gaps. The Nilsson-Ehle Endowments (43255) and the Lundman’s Foundation for Botanical Studies grants from the Swedish Phytogeographic Society awarded to Qiujie Zhou covered the cost of high-throughput sequencing.
CONFLICT OF INTEREST STATEMENT
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are openly available in "NCBI sequence read archive" at https://www.ncbi.nlm.nih.gov/sra, all raw reads under PRJNA511374 and PRJNA1007582. The scripts used for all the analyses, simulations, and respective input data are openly available in GitHub repository, at https://github.com/piyalkarunam/RC_EAA (DOI: https://doi.org/10.5281/zenodo.10809862).

ORCID
Piyal Karunarathne https://orcid.org/0000-0002-1934-145X
Qiujie Zhou https://orcid.org/0000-0001-7351-2371
Martin Lascoux https://orcid.org/0000-0003-1699-9042
Pascal Milesi https://orcid.org/0000-0001-8580-4291

REFERENCES


**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.

---

**How to cite this article:** Karunarathne, P., Zhou, Q., Lascoux, M., & Milesi, P. (2024). Hybridization mediated range expansion and climate change resilience in two keystone tree species of boreal forests. *Global Change Biology*, 30, e17262. [https://doi.org/10.1111/gcb.17262](https://doi.org/10.1111/gcb.17262)