Patterns of genotype-environment association in the eastern North American yellow birch (*Betula alleghaniensis* Britt.)

By

Myles Cummins

A graduate thesis submitted in partial fulfillment of the requirements

for the degree of Master of Science in Forestry

Faculty of Natural Resources Management

Lakehead University

ABSTRACT

Cummins, M (2025), Patterns of genotype-environment association in the eastern North American yellow birch (*Betula alleghaniensis* Britt.), Master of Science in Forestry, Lakehead University. Advisor: Dr. A.M. Thomson.

Understanding how genomic adaptation shapes species' responses to climate change is essential for developing climate-resilient forests, as shifting conditions increasingly drive range shifts and maladaptation. This study investigates adaptive genomic variation in *Betula alleghaniensis* (yellow birch), a widely distributed hardwood of eastern North America. Genome-wide single-nucleotide polymorphism (SNPs) variation from 27 populations was analyzed using 3D-genotype-by-sequencing and two genotype-environment association methods: redundancy analysis (a multivariate ordination method) and Gradient Forest (a machine learning algorithm). 124 putatively adaptive loci were identified, linked to extreme minimum temperature, degree-days below 0°C, winter precipitation, and snowfall. Functional annotation revealed roles primarily in stress response and transcriptional regulation. Patterns of adaptive variation showed a latitudinal gradient tied to winter severity and spatially heterogeneous responses to snowfall. Two distinct clusters of adaptive loci were identified along the climate gradients, suggesting winter climate plays a dominant role in shaping local adaptation. Future climate projections (SSP5-8.5, 2041-2070) predict substantial shifts in adaptive alleles in the Northeastern Appalachians, Maritimes, and St. Lawrence River regions. Nevertheless, genetic offset, the Euclidean distance between the current and future adaptive genomic composition, across the range was relatively low, suggesting genomic resilience potentially supported by yellow birch's allohexaploid genome and extensive gene flow, including adaptive introgression from hybridization with other *Betula* members. These findings highlight the importance of integrating genomic data into forest management strategies.

Keywords: *Betula alleghaniensis*, Yellow Birch, local genetic adaptation, genotype-environment association (GEA), genetic offset, climate change

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Ashley Thomson, for her guidance and support throughout my graduate thesis. Her expertise and feedback helped positively shape the direction of my work, and I deeply appreciate the time and effort they contributed to both my academic and career development. I am grateful to my committee members, Dr. Nathan Basiliko and Dr. Jian Wang, for their insights and guidance throughout my thesis. The scope of this work would not have been possible without sample contributions from Dr. Thomson, Dr. Patrick Lenz (NRCAN), Cameron Cornelson and the National Tree Seed Centre. I appreciate the lab support and resources provided by Dr. Mike Rennie, Dr. Basilko, Dr. Susanne Walford, and Mike Moore throughout the long months of extracting more tears than DNA. Thank you to the Lakehead University Greenhouse manager, Keri Pidgen-Welyki, for providing me with a place to escape the frigid winter of lab work for fun, friendship and fantastic flora, as well as providing space for seedling germination. Special thanks to my lab mates Karla Ramirez Galicia, Hailey Orchard, and Hannah Malloy, who took time from their research to help me stay motivated in the lab amidst the ever-growing mountains of microcentrifuge tubes. I also want to thank Yuanxin Ye for their invaluable technical support in performing gene ontology. Finally, I am immensely grateful to the National Science and Engineering Research Council (NSERC) for funding this research.

Outside the academic roots of this project, I am deeply thankful to the people who helped me stay grounded and kept me growing, even when things felt wilted. Thank you to Dr. Emma Lehmberg for her mentorship and support throughout this endeavour, and for listening to my endless catharsis; your kindness helped me stay rooted when things got a little overgrown. I am deeply grateful to the two people who first sowed the seeds of my passion for botany, my mother Jill Skeoch-Cummins and our neighbour Bill Baulch; their early encouragement helped this journey take root. I also want to thank those who nurtured my growing interest in academia: Dr. Lada Malek, Dr. Peter Lee, and Dr. Stephen Hecnar, whose teaching and mentorship helped my path continue to flourish. And my dear friend, Kristi Valley, for her constant friendship and years of encouragement, perspective, and laughter. From early days among pressed plants to conversations that helped me clarify far more than plant IDs, your support has meant the world to me.

Last, I would like to give a heartfelt thanks to my partner, Ashley Levins, whose love, caring, and unwavering support made obtaining this degree and all that came with it possible. Thank you for believing in me on the days I couldn’t, for celebrating every small win, and for reminding me there’s life beyond the research. Your presence has made this journey more meaningful, and I’m endlessly blessed to walk through this chapter and the next with you. And of course my cats, Conky and Pepé, for keeping me sane, offering distractions both ill-timed and oddly essential, and making even the longest coding days bearable.

Table of contents

LIST OF TABLES

LIST OF FIGURES

LIST OF SUPPLEMENTAL MATERIAL

Supplemental Methods

Supplemental Figures

Supplemental Tables

INTRODUCTION

Climate change is rapidly reshaping forest ecosystems, altering temperature and precipitation regimes, with direct consequences for tree growth and survival (Allen *et al.*, 2010). Forest tree populations must respond to these rapidly changing conditions through migration, phenotypic plasticity, or genetic adaptation. However, the rate of environmental change is expected to exceed the capacity of many species to adapt through standing genetic variation (Aitken *et al.*, 2008), potentially leading to widespread maladaptation to climate. These changes can have far-reaching consequences for forest biodiversity, ecosystem services, and the economic value of timber-producing species (Liang *et al.*, 2016; Weiskopf *et al.*, 2020). To mitigate these risks, a deeper understanding of local adaptation and genetic diversity patterns across species ranges is needed to inform conservation strategies like assisted migration and gene flow (Aitken & Whitlock, 2013; Park *et al.*, 2018).

Broadly distributed tree species often occupy diverse environmental gradients and exhibit population-level genetic differentiation shaped by local climate (Savolainen *et al.*, 2007). Traditional provenance trials have been foundational in studying adaptive trait variation across climate gradients (Sork *et al.*, 2013), offering insights into growth, phenology, and stress responses. Yet such trials are limited in number and geographic scope, leaving many ecologically significant species underrepresented. Advances in next-generation sequencing now allow detection of adaptive genetic variation even in species lacking common-garden experiments. Genotype-environment association (GEA) methods, which examine correlations between allele frequencies and environmental variables, have become powerful tools for identifying local adaptation and detecting loci under selection (Holliday *et al.*, 2017; Isabel *et al.*, 2020; Yu *et al.*, 2022).

Multivariate approaches such as redundancy analysis and machine-learning tools like gradient forests are especially useful in detecting both linear and nonlinear responses to environmental gradients (Capblancq *et al.*, 2020; Rellstab *et al.*, 2016). These tools can also identify subtle allele-environment correlations that might otherwise go undetected, particularly in polygenic traits (Forester *et al.*, 2018; Láruson *et al.*, 2022). They also enable landscape-level predictions of adaptive genetic variation by integrating spatial environmental heterogeneity (Forester *et al.*, 2016). Importantly, GEA models can incorporate demographic history to account for population structure and historic gene flow, which are crucial for accurately identifying loci under selection (Lotterhos & Whitlock, 2015; Tibbs Cortes *et al.*, 2021). Importantly, they can predict spatial patterns of adaptive genomic variation and forecast how genotype–environment associations may shift under future climate scenarios (Fitzpatrick & Keller, 2015; Sang *et al.*, 2022). This predictive capacity is particularly valuable for species lacking comprehensive experimental trials and can inform conservation planning across heterogeneous landscapes.

In addition to identifying adaptive loci, GEA-based methods can estimate the genetic offset, the expected disruption of genotype-environment associations under projected climate change (Fitzpatrick & Keller, 2015). Genetic offset quantifies the magnitude of allele frequency change required for populations to remain locally adapted, providing a spatially explicit measure of vulnerability (Rellstab *et al.*, 2021). These models help identify both vulnerable populations and potential climate refugia, offering valuable guidance for seed sourcing and assisted gene flow (Fitzpatrick & Keller, 2015; Meek *et al.*, 2023) and have been applied to numerous tree species, including *Araucaria araucana* (Varas-Myrik *et al.*, 2024), *Picea rubens* (Lachmuth *et al.*, 2023) and *Plathymenia* spp. (Muniz *et al.*, 2024).

Yellow birch (*Betula alleghaniensis* Britt.), a hexaploid species (2n=6x=84; Ashburner *et al.*, 2013), is a broadly distributed hardwood species of the mesophytic mixed forests of eastern North America (Erdmann, 1990). It is ecologically significant, providing a vital food and nesting source for a wide range of wildlife species (Holmes & Schultz, 1988; Menzel *et al.*, 2004), and holds commercial value for its hardwood lumber (Erdmann, 1990). Yellow birch is typically associated with late-successional forests but occurs in a variety of successional stages and site conditions, demonstrating a degree of ecological flexibility. It plays an integral role in nutrient cycling and forest structure, influencing community dynamics throughout its range. However, climate envelope models project a reduction in abundance and biomass under future warming scenarios (Van Houtven *et al.*, 2019), suggesting potential ecological and economic consequences.

Previous provenance trials of yellow birch revealed clinal variation in adaptive traits across climate gradients (Clausen, 1968a, 1968b, 1975, 1977, 1980; Danick & Barnes 1975; Sharik and Barnes, 1975, 1979; Wearstler and Barnes, 1977; Leites *et al.*, 2019; Pedlar *et al.*, 2021; Maloney *et al.*¸2024). Clausen (1968b) documented clinal variation in growth cessation correlated with latitude, while Wearstler and Barnes (1977) linked total height and seed germination to growing season length. Recent studies have shown climate-associated variation in height growth (Leites *et al.*, 2019) and leaf morphology (Maloney, 2022), highlighting environmental influence on genetic differentiation. However, these studies also point to substantial within-population genetic variation, suggesting a complex interplay of local adaptation and genetic diversity. This within-population diversity may buffer yellow birch against environmental change, although the extent to which it will be sufficient under rapid climate change remains unclear.

At the genomic level, yellow birch's hexaploid nature may enhance its adaptive capacity by providing genetic redundancy and promoting greater allelic diversity (Soltis & Soltis, 2009;

Rice *et al.*, 2019). Polyploidy in plants is often associated with increased evolutionary potential and resilience to environmental stress, facilitating adaptation to heterogeneous and changing conditions (Lavania, 2020). Furthermore, historic hybridization events between yellow birch and related species, *Betula papyrifera* and *B. pumila* (Thomson *et al.*, 2015; Wang *et al.*, 2016), have likely contributed to its genetic variation. Introgression from these species may have introduced adaptive alleles that facilitate resilience to climatic shifts, a phenomenon increasingly recognized as a driver of adaptive evolution in tree taxa (Taylor & Larson, 2019). These genomic features suggest that yellow birch may be particularly well equipped to respond to environmental change through adaptive evolution.

This study investigates the genomic basis of local adaptation in yellow birch across its climatic range. Given its broad distribution, high genetic diversity, polyploid genome, and history of hybridization, yellow Birch is expected to exhibit significant climate-related genetic differentiation. Yet, the spatial patterns of this adaptation and the potential for future maladaptation remain poorly understood. To address this gap, genome-wide allele frequency data and genotype-environment association (GEA) were used to characterize patterns of local adaptation. Specifically, the study aimed to: (i) identify the climatic drivers of genetic differentiation, (ii) detect putatively adaptive loci associated with climate gradients, (iii) assess the genetic offset required for populations to maintain local adaptation under a high-emissions climate change scenario. Integration of redundancy and Gradient Forest analysis enabled evaluation of the spatial distribution of adaptive genetic variation and population-level vulnerability to climate change. The findings provide a genomic framework to inform conservation planning, seed transfer guidelines, and assisted gene flow strategies aimed at sustaining yellow birch populations in a changing climate.

MATERIALS & METHODS

Sample collection and DNA extraction

A total of 27 populations of yellow birch were sampled across its eastern North American distribution (Figure 1; Supplemental Table 2). This included silica-dried leaf tissue obtained from a prior study (Thomson, 2013) and newly collected material gathered in 2023. Yellow birch seed was also acquired from the National Tree Seed Centre (NTSC). Seeds were cold stratified, germinated for three weeks in the Lakehead University greenhouse, and grown for 2-3 weeks before being frozen at -20°C.

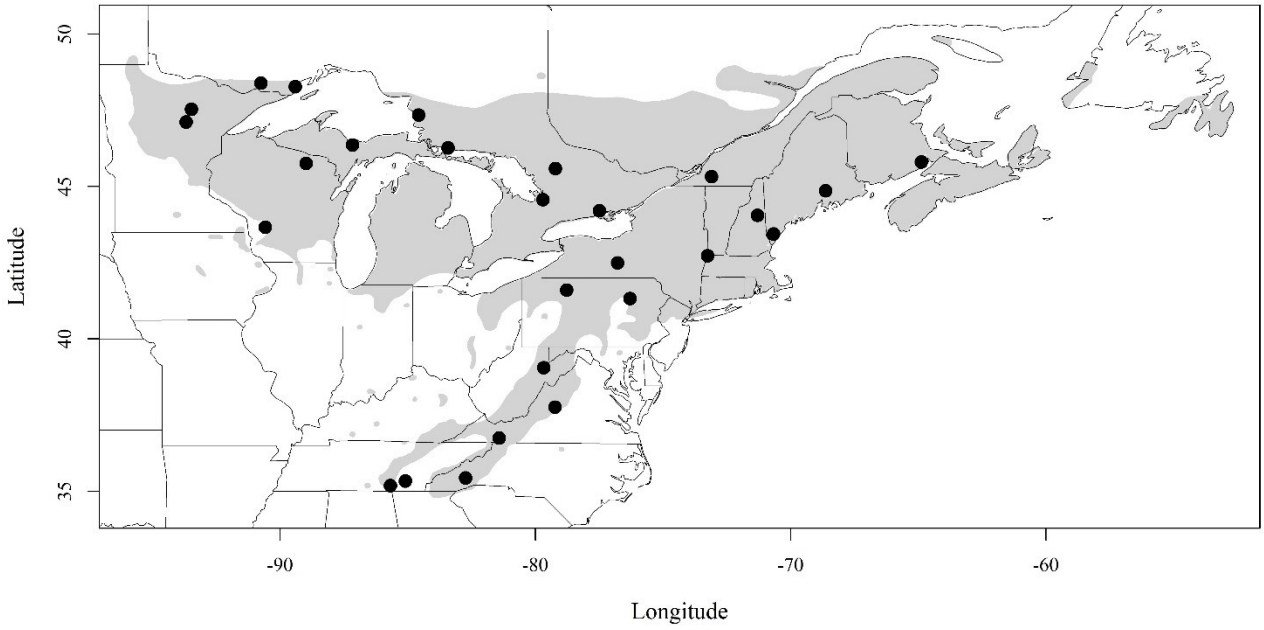


Figure 1. Sampling locations for 27 *Betula alleghaniensis* populations across its eastern North American distribution. The known species distribution is shown based on range maps from Little (1971).

Genomic DNA was extracted from 5-10 individuals per population using either 20 mg of silica-dried leaf tissue or whole seedlings. A modified 3% CTAB protocol adapted from Zeng *et al.* (2002) was applied to both tissue types. DNA integrity and purity were assessed using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and Qubit 4 Fluorometer with a dsDNA Broad Range Assay (Thermo Fisher Scientific, Waltham, MA, USA). Samples failing to meet the concentration threshold of 20 ng/µL or containing high levels of contaminants were re-extracted. Full extraction protocol details are provided in Supplementary Methods S1.

GBS library preparation and bioinformatics processing

A total of 243 yellow birch samples were diluted to 10 ng/µL and submitted to the Institut de Biologie Intégrative et des Systèmes (IBIS) at Université Laval for library preparation using the 3D-genotype-by-sequencing (3D-GBS) protocol (de Ronne *et al.*, 2020). This protocol employs a three-enzyme digestion (*PstI*, *NsiI*, and *MstI*) designed to enrich for gene-rich, low-redundancy regions while reducing repetitive content. The method is particularly well suited for polyploid genomes, including hexaploids, as it improves sequencing depth and SNP-calling accuracy while minimizing allele dropout due to complex allele dosage and redundancy.

Libraries were sequenced at the Genome Québec Innovation Centre in Montréal using an Illumina NovaSeq platform to generate 35 million, 150 bp, paired-end reads. IBIS performed read trimming, alignment to the *Betula pendula* reference genome (GenBank GCA_900184695.1), and SNP calling with a STACKS pipeline (Catchen *et al., 2013;* Normandeau, n.d.). After quality control, adapter trimming, demultiplexing, and multi-step SNP filtering, a total of 32,295 high-

confidence SNPs were retained across 221 individuals. Detailed steps and filtering parameters are provided in Supplementary Methods S2.

To account for the hexaploid nature of yellow birch and reduce bias from potential over-assembly of paralogous loci, an additional filtering step was implemented using the polyRAD R package (Clark *et al.*, 2019, 2022). Loci with high overdispersion or excess heterozygosity ($H_{ind}/H_e > 0.833$) were excluded, resulting in a final set of 29,674 loci. Population structure analysis was conducted using principal component analysis (PCA), and three putative hybrid individuals were identified and removed, leaving 218 individuals for downstream analysis. The first two PCA axes were averaged at the population level and used as covariates in subsequent analyses. Complete procedures are provided in Supplementary Methods S3.

Selection of climate variables

Climate variables were obtained from ClimateNA for the 1961-1990 normals (Wang *et al.*, 2016) covering the range of yellow birch. A Pearson's pairwise correlation analysis was used to remove highly collinear variables ($r \geq 0.7$), resulting in the retention of nine uncorrelated predictors: annual heat-moisture index (AHM), degree-days below 0°C (DD_0), degree-days above 5°C (DD5), extreme minimum temperature (EMT), mean annual solar radiation (MAR), precipitation as snow (PAS), summer precipitation (PPT_sm), winter precipitation (PPT_wt), and mean annual relative humidity (RH). These nine climate variables were retained for all subsequent analyses involving environmental predictors throughout the study.

Partitioning genetic variance

To evaluate the relative contributions of environmental, geographic, and neutral processes to genetic variation across the range of yellow birch, a series of redundancy analyses (RDA) and partial redundancy analyses (pRDA) were performed following Capblancq et al. (2023). RDA is a multivariate ordination approach that models linear relationships between genetic variation and predictor variables, allowing for the isolation of independent (or "pure") effects by conditioning on potential confounders such as geographic structure or demographic history (Capblancq & Forester, 2021). Local allele frequencies were used as multivariate response variables in all models.

Geographic structure was modelled using the first three distance-based Moran's eigenvector maps (dbMEM) axes derived from the 27 sample population coordinates. To account for neutral population structure, the mean PC1 and PC2 axes scores per population from the polyRAD principal component analysis were included as proxies for background genetic structure. Climate was represented by the nine selected climate variables. A full RDA was constructed using all predictor sets (9 climate variables, 3 dbMEM axes, and PC1 and PC2), and three pRDAs were run to isolate the unique effects of climate, geography, and population structure. Full modelling procedures are detailed in Supplementary Methods S4.

Identification of putatively adaptive loci

To identify loci strongly associated with climate variation, two complementary genotype-environment association (GEA) approaches were employed: redundancy analysis (RDA) and Gradient Forests (GF). RDA's uses in GEA's has been shown to be particularly effective for

detecting subtle, polygenic signals of adaptation by identifying multivariate correlations between allele frequencies and environmental gradients while accounting for confounding structure (Forester *et al.*, 2018). In contrast, GF is a machine-learning algorithm that models non-linear relationships and variable interactions and is well suited for capturing complex genotype-environment associations (Ellis *et al.*, 2012). When used in tandem, these methods provide a robust insight into local adaptation by balancing the strengths and limitations of each (Fitzpatrick & Keller, 2015; Capblancq *et al.*, 2023).

Given the potential confounding influence of neutral genetic structure, both methods were run in raw (RDA-raw, GF-raw) and structure-corrected (RDA-x, GF-x) formats. While accounting for neutral structure can reduce false positives, it may also obscure true climate-driven signals of adaptation, especially in species where population structure, geography, and environment are strongly collinear (Capblancq *et al.*, 2023). RDA-x models were conditioned on PC1 and PC2 from the polyRAD population structure PCA. For GF-x allele frequencies were standardized using a Bayesian multivariate approach implemented in JAGS (Plummer, 2003, 2024), where the covariance matrix was sampled via MCMC and Cholesky decomposition was applied to standardize allele frequencies, following the framework of Günther & Coop (2013).

SNPs were ranked within each GEA test based on Mahalanobis distance (RDA-raw and RDA-x) or R^2 scores (GF-raw and GF-x). From each test, the top 100 ranked SNPs were retained. Loci identified among the top 100 in at least two of the four tests were considered putatively adaptive. This consensus approach helps minimize false positives while capturing loci consistently associated with environmental gradients across both linear and non-linear models. The resulting set of adaptively enriched loci were used to construct composite adaptive indices. Complete detection, standardizing, and ranking procedures are provided in Supplementary Methods S5.

Modelling the spatial distribution of adaptive genomic variation

To model how adaptive genomic variation is distributed across the range of yellow birch, an RDA was performed using the set of putatively adaptive SNPs and the nine selected climate variables. The loadings of these climate predictors on the first two RDA axes were combined with standardized climate data to construct two composite adaptive indices (RDA1 and RDA2), representing major axes of climate-associated genetic variation (Steane *et al.*, 2014; Capblancq *et al.*, 2020). These indices were projected across the current distribution of yellow birch at 1-km spatial resolution using a standardized Climate NA 1961-1990 normal dataset.

To explore whether subsets of adaptive loci were associated with specific environmental gradients, k-means clustering was applied to the absolute values of SNP loadings along the first two RDA axes. Clusters were defined using the absolute values of adaptive locus scores on RDA1 and RDA2, each representing loci primarily associated with one RDA axis, characterized by a high absolute score on one axis and a low score on the other.

To forecast future genomic-climate relationships, projections for the nine climate variables were obtained from ClimateNA under the 8GCM SSP5-8.5 emissions scenario (heavy fossil-fuel development scenario) for the 2041-2070 period (Mahony *et al.*, 2022). Future adaptive indices for both RDA1 and RDA2 were computed using the same RDA loadings and standardized climate projections, enabling spatial predictions of genomic adaptation under anticipated climate change. Full details on the index construction, clustering approach, and mapping are provided in Supplementary Methods S6.

Genetic offset and adaptive capacity

To estimate the genomic change required for populations of yellow birch to remain adapted under future climate scenarios, the genetic offset was calculated across the range. This measure represents the Euclidean distance between present and future adaptive indices derived from RDA1 and RDA2 (Capblancq *et al.*, 2020). Current indices were subtracted from future projections, and the resulting differences across RDA1 and RDA2 were summed to generate a single genetic offset value for each 1-km^2 cell across the distribution of *yellow* birch. This value reflects the degree to which populations would need to shift adaptive allele frequencies to maintain climatic optima and serves as a proxy for vulnerability to climate change (Fitzpatrick & Keller, 2015).

To assess the evolutionary capacity of populations to respond to such changes, two metrics were calculated: standing genetic variation (SGV) and the population adaptive index (PAI). SGV quantifies the within-population variance in adaptive allele frequencies, estimated as the mean value of p x q at each adaptive SNP, reflecting the availability of adaptive alleles within each population (Chhatre *et al.*, 2019). PAI measures the degree to which a population's adaptive allele frequencies deviate from the species-wide mean, calculated as the absolute difference from the average allele frequency across all populations (Bonin *et al.*, 2007). While SGV reflects a population's capacity for adaptation by quantifying available variation at adaptive loci, PAI measures the extremeness of that capacity by capturing how far a population's adaptive genetic composition deviates from the species-wide average. Full computation procedures are detailed in Supplementary Methods S7.

Gene function annotation

To infer the potential functional roles of adaptive loci, SNPs were annotated using a draft *Betula pendula* genome (GenBank GCA_900184695.1). Gene prediction was performed using AUGUSTUS v3.5.0 (Stanke *et al.*, 2006) with the *Arabidopsis* gene model. SNPs were intersected with predicted gene regions using BEDTools v2.30.0 (Quinlan & Hall, 2010), and nearby genes were assigned to intergenic variants. Functional annotations were assigned through orthology-based prediction using eggNOG-mapper v2 (Cantalapiedra *et al.*, 2021). This approach enabled inference of the putative functions of climate-associated genomic regions in yellow birch. Complete annotation procedures are described in Supplementary Methods S8.

Code development and data visualization

Portions of the data analysis were assisted using OpenAI (GPT-4, 2025). The model was used iteratively to support code development, data wrangling, and the production of figures for exploratory and final visualization. All code was reviewed and executed by the lead author.

RESULTS

Drivers of genetic variation in yellow birch

After filtering and paralog removal using polyRAD, 29,674 high-quality loci were retained for downstream analyses of genetic variation. Local allele frequencies at these loci formed the response variables in redundancy analyses (RDA and partial RDA), which revealed that a substantial portion of genetic variation in yellow birch was explained by climate, geography, and neutral population structure. The full model, which included all three groups of predictor variables, accounted for 64% of the total variance (Table 1; Supplemental Figure S1). The climate-only model, while not significant after controlling for geographic and neutral population structure, explained 29% of the total variance (equivalent to 46% of the explainable variance). Geographic structure similarly was not significant when controlling for climate and neutral population structure but still accounted for 10% of the total variance (16% of the explainable variance). In contrast, the neutral population structure model was significant even when controlling for climate and geography, explaining 9% of the total variance (14% of the explainable variance). A moderate level of shared (confounded) variance among the three predictor groups accounted for 15% of the total variance (24% of the explainable variance).

Table 1. Summary of pRDAs used to partition the genetic variance in yellow birch among climate (clim.; 9 select climate variables), geographic structure (geog.; 3 dbMEM axes), and neutral population structure (ancestry, struc.; 2 mean PC axes). Each set of predictors was tested individually and jointly. This presents the percentage of total genetic variance explained, the significance of the model, and the proportion of variance attributed to the full set of explanatory variables.

RDA Models	**Inertia**	**P(>F)**	**Proportion of explainable inertia**	**Proportion of total inertia**
Full model: F ~ clim. + struc. + geog.	24.4	0.001***	1.00	0.64 (R^2)
Pure climate: F ~ clim. | (geog. + struc.)	11.1	0.197	0.46	0.29 (R^2)
Pure geography: F ~ geog. | (clim. + struc.)	3.9	0.252	0.16	0.10 (R^2)
Pure ancestry: F ~ struc. | (clim. + geog.)	3.4	0.027*	0.14	0.09 (R^2)
Confounded climate/geography/structure	5.9		0.24	0.15
Total unexplained	13.8			0.36
Total inertia	38.1			1.00

Genotype-environment associations

Many of the top 100 loci were shared across the four multivariate tests, resulting in a total of 263 unique top-ranking loci (Figure 2). The greatest degree of overlap occurred between the top SNPs from the same test performed using raw allele frequencies and with correction for population structure. The two RDA tests showed the highest congruence, with 84% of loci shared between them. Similarly, the top loci from the two GF tests exhibited a high degree of overlap (44%). In contrast, overlap between RDA and GF tests was low, regardless of whether population structure was corrected for, with only 6-7% of the top SNPs shared between these test types. Overall, 124 loci were identified in the top 100 SNPs of at least two tests, and five loci were consistently detected by all four GEA tests.

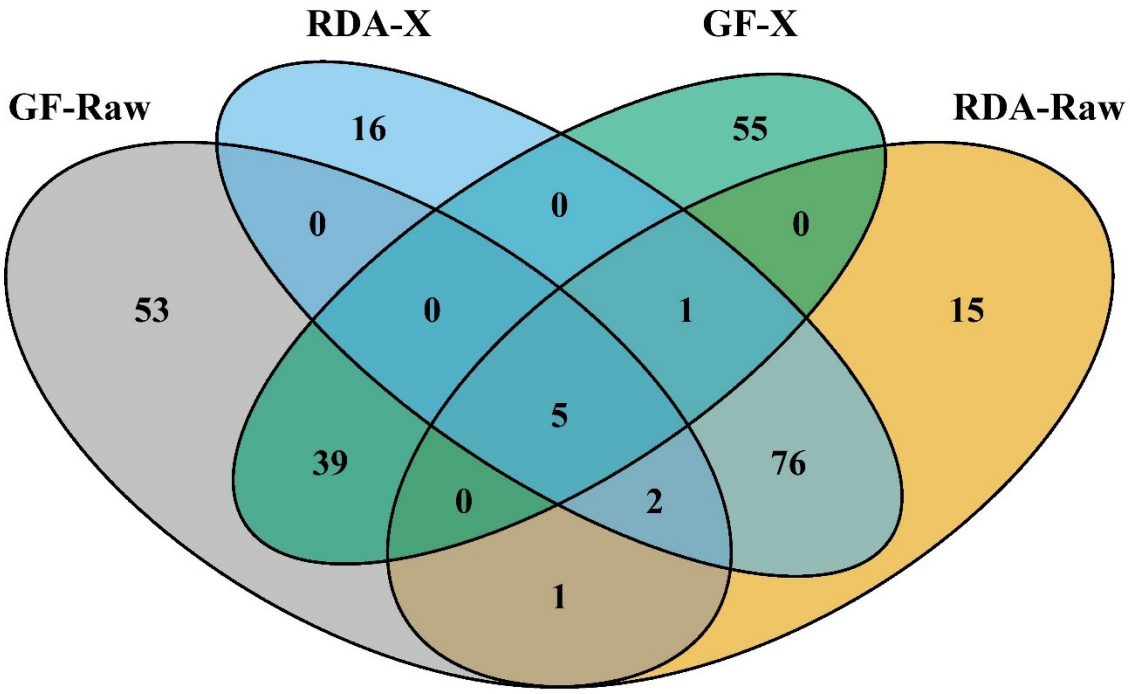


Figure 2. Venn diagram showing the overlap of the top 100 outlier loci identified by each of the four GEA methods, with each ellipse representing a different detection method; RDA-raw: redundancy analysis using mean allele frequencies, RDA-X: redundancy analysis using mean allele frequencies conditioned on neutral genetic structure, GF-raw: Gradient Forest analysis using mean allele frequencies, GF-X: Gradient Forest analysis using standardized mean allele frequencies.

Climate associations of adaptive loci

The adaptively enriched RDA (conducted on the 124 putatively adaptive loci identified by the GEA tests) revealed that the first two constrained axes explained 61.2% of the total variance in allele frequencies. RDA1 accounted for 34.5% of this variance and represented a composite environmental gradient strongly positively correlated with extreme minimum temperature (0.92) and winter precipitation (0.91), and strongly negatively associated with degree-days below 0°C (-

0.91) (Figure 3a). RDA2 explained 11.5% of the variance and was most strongly associated with precipitation as snow (-0.42).

The adaptively enriched RDA also identified clusters of loci associated with the two primary climate gradients captured by RDA1 and RDA2 (Figure 3b; Supplemental Figure S2). Two distinct clusters were strongly associated with a single axis, and a group of nine loci showed strong correlations with both. The first cluster, linked to RDA1, comprised 25 adaptive loci whose mean allele frequencies showed a clear clinal trend, increasing with colder extreme minimum temperatures across populations ($r^2 = 0.48$; Figure 3c). The second cluster, associated with RDA2, included 33 loci with only a weak clinal relationship with precipitation as snow ($r^2 = 0.13$; Figure 3d). Notably, a southern Appalachian population at the extreme southern range margin exhibited unexpectedly high allele frequencies for the RDA2 cluster.

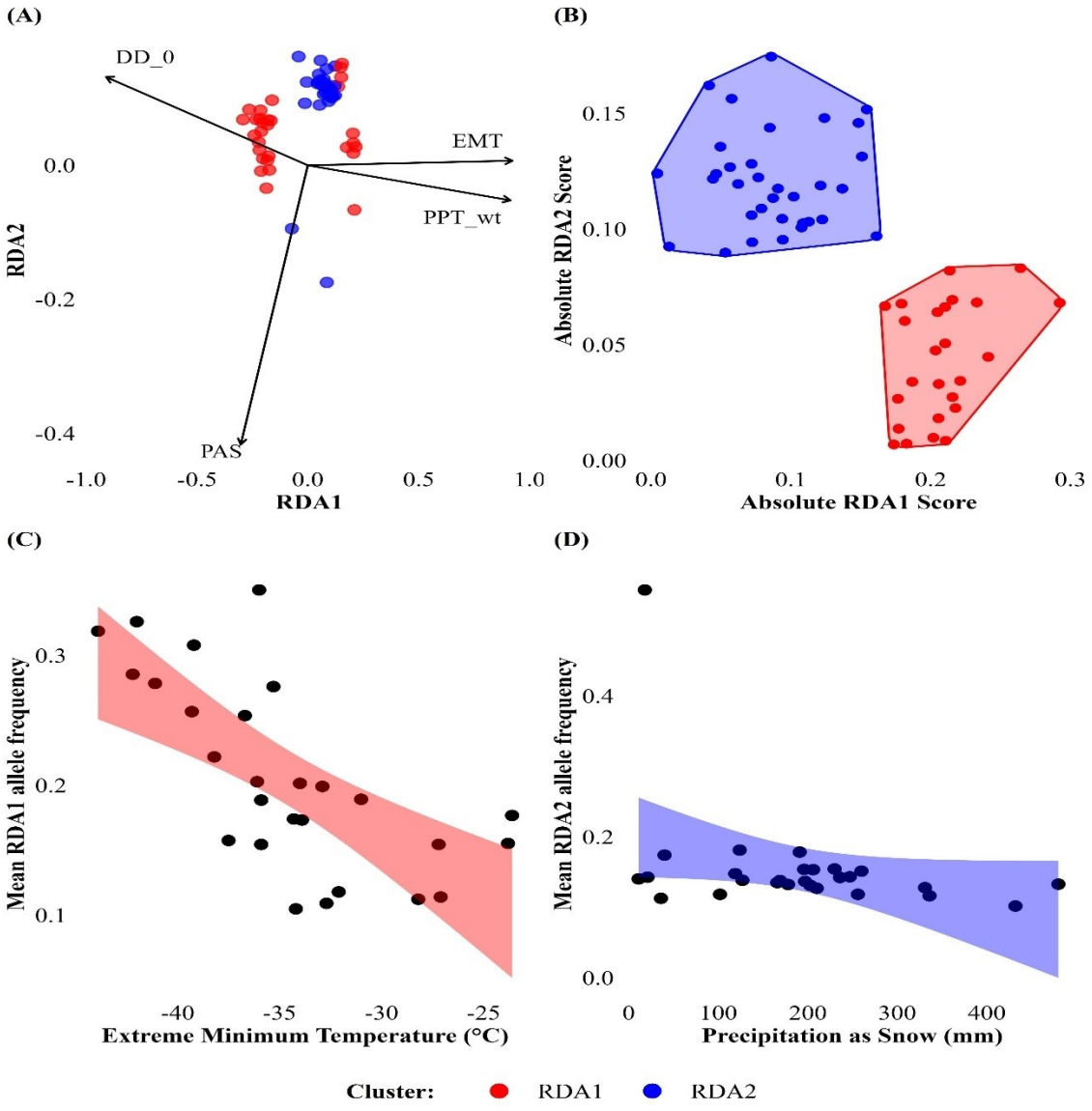


Figure 3. Red and blue colours indicate SNPs associated with RDA1 and RDA2, respectively, and are used consistently across all four panels. **(A)** Biplot of SNP loadings on the first two axes of the adaptively enriched RDA, showing loci (points) colored by their strongest axis of association (RDA1 or RDA2). Arrows represent the four most strongly associated climate variables: extreme minimum temperature (EMT), degree-days below 0°C (DD_0), winter precipitation (PPT_wt), and precipitation as snow (PAS). **(B)** Absolute RDA1 and RDA2 scores of adaptive loci, grouped by primary axis of association. Convex hulls outline the clusters of loci strongly associated with each RDA axis. **(C)** Relationship between mean allele frequencies of the 25 RDA1-associated loci and extreme minimum temperature across 27 yellow birch populations. Shaded areas indicate the 95% confidence interval around the linear regression. **(D)** Relationship between mean allele frequencies of the 33 RDA2-associated loci and precipitation as snow across the same populations, including the 95% confidence interval. between mean 33 RDA2-associated allele frequencies and precipitation as snow (mm) across 27 sampled populations, with a 95% confidence interval along

the linear regression. Frequencies shown in **C** and **D** reflect mean non-reference alleles and reflect loci with both positive and negative associations along each RDA axis.

Spatial projections of contemporary and future adaptive landscapes

Mapped patterns of adaptive genetic variation, derived from the climate gradients captured by the first two axes of the adaptively enriched RDA, revealed distinct spatial structure across the range of yellow birch (Figure 4a). The RDA1-based adaptive index displayed a strong latitudinal gradient, with lower values corresponding to adaptation to longer, colder, and dryer winters in the northern part of the range. In contrast, the RDA-2-based index reflected a combination of latitudinal and longitudinal structure. Lower RDA2 values, associated with adaptation to higher snowfall, were concentrated in the Maritime provinces, the Laurentian Mountains, and high-elevation regions of the northeastern United States (Figure 4b).

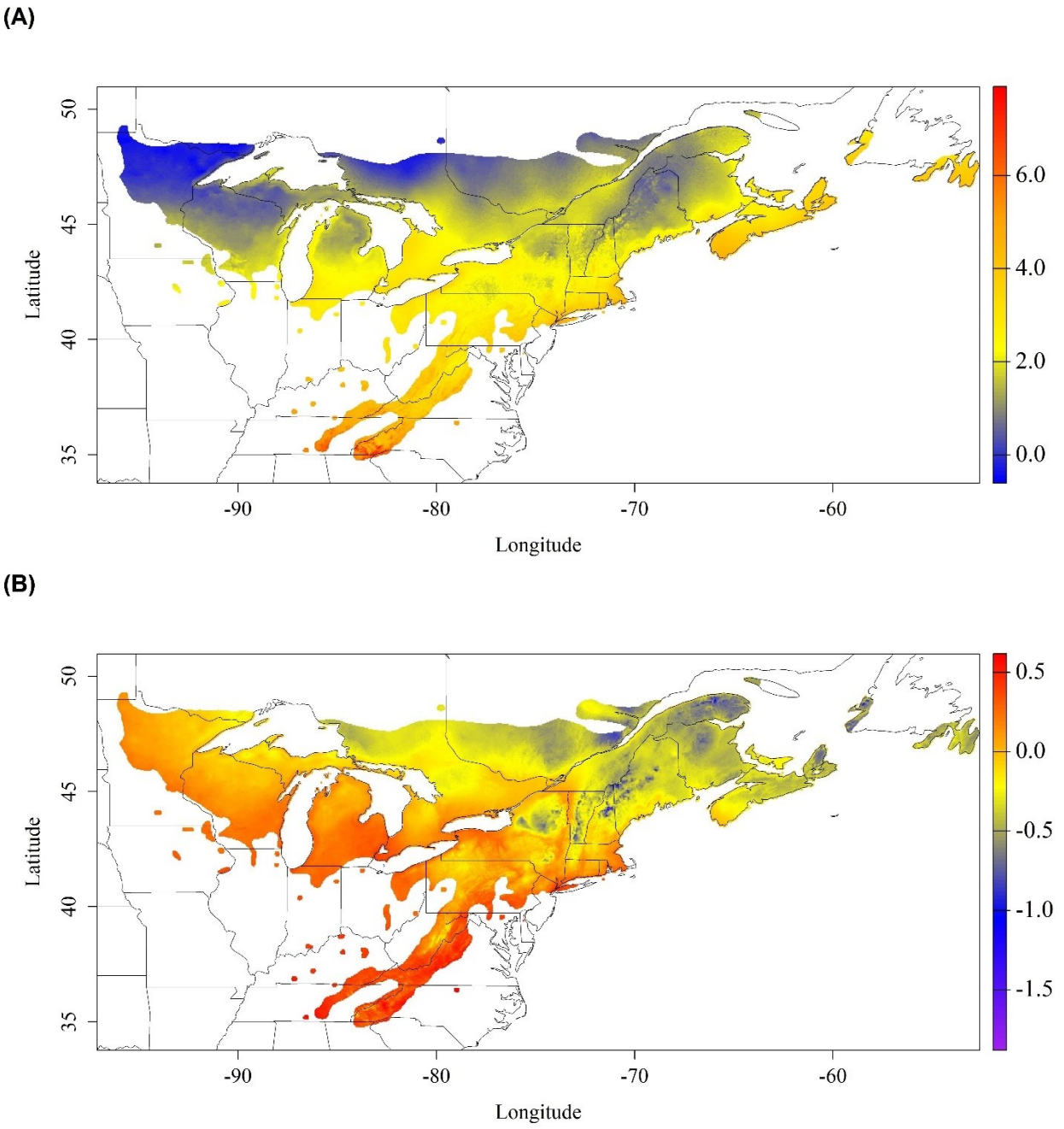


Figure 4. Spatial projections of adaptive indices derived from nine climate variable loadings produced from the adaptively enriched RDA and 1960-1991 climate data. Colour gradients represent a dimensionless score calculated by RDA loadings and standardized climate variables. **(A)** RDA1 contemporary adaptive landscape; **(B)** RDA2 contemporary adaptive landscape.

Future projections under a high-emissions scenario (2041-2070) suggest that the geographic distribution of these adaptive patterns may shift. While the overall spatial structure of both RDA axes remained broadly similar to present-day patterns, notable changes emerged, particularly for RDA1. The projected range of index values for RDA1 increased, suggesting more pronounced differences in adaptation to winter conditions across the landscape (Figure 5a). These shifts were concentrated in the northern Appalachian Mountains, the St. Lawrence River valley, and parts of the central and eastern Great Lakes region. For RDA2, projected increases in index values suggest a growing mismatch between current snowfall-associated genetic adaptation and future environmental conditions in the Maritimes, the northern Appalachians, southern Ontario, and Michigan (Figure 5b).

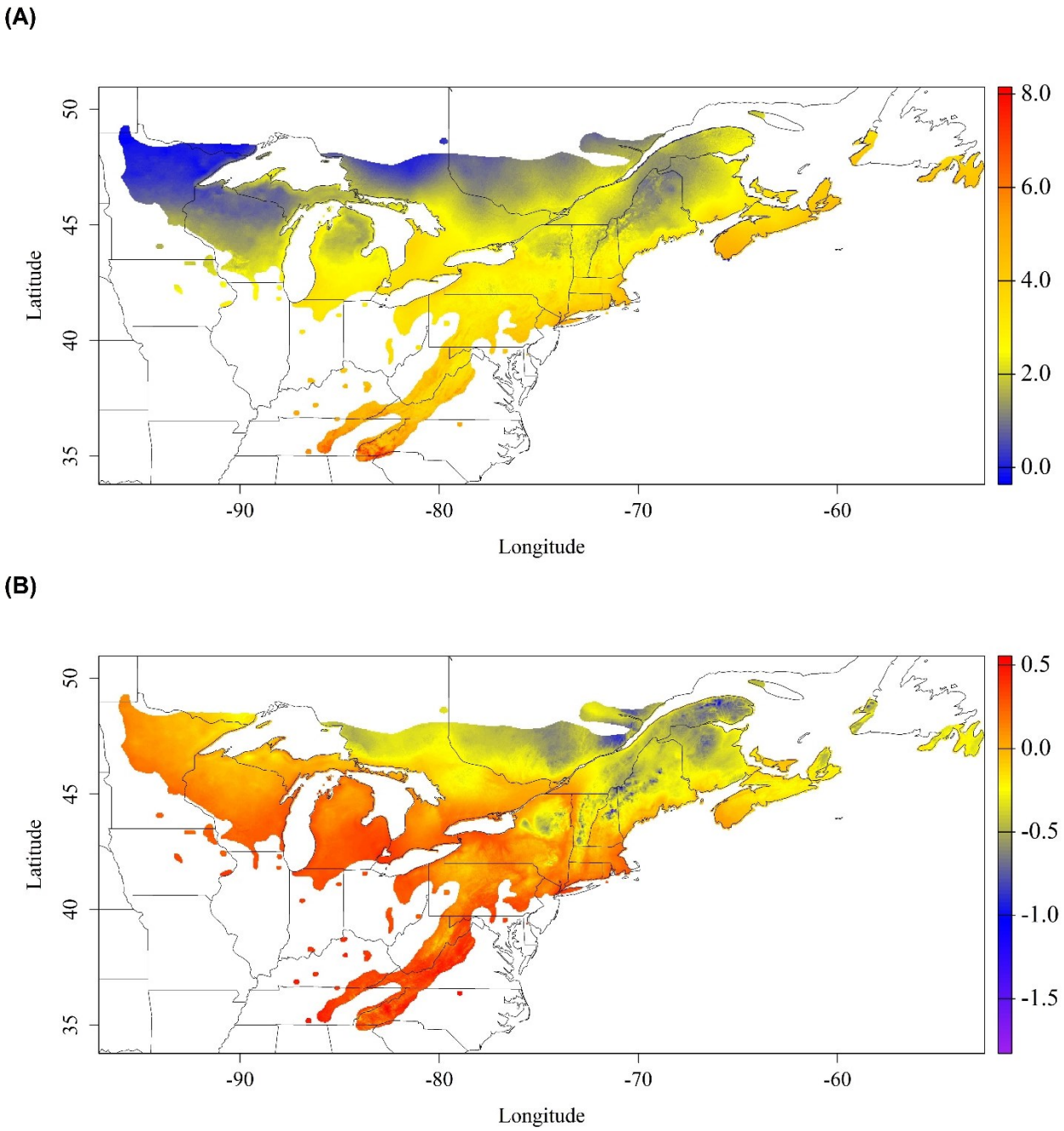


Figure 5. Spatial projections of adaptive indices derived from climate variable loadings produced from the adaptively enriched RDA and SSP5-8.5 2041-2070 climate data. Colour gradients represent a dimensionless score calculated by RDA loadings and standardized future climate variables. **(A)** RDA1 future adaptive landscape; **(B)** RDA2 future adaptive landscape.

Genetic offset & adaptive capacity of yellow birch populations

The genetic offset revealed a broad longitudinal pattern of increasing mismatch between current and future adaptive variation moving eastward across the range of yellow birch (Figure 6). Offset values were highest in the Maritimes and northern parts of the Northeastern Appalachian Mountains, indicating a greater degree of change required for populations to remain adapted under future climate conditions. In contrast, lower offset values were observed in the western Great Lakes and the southern Appalachian Mountains, suggesting these regions may face less adaptive disruption.

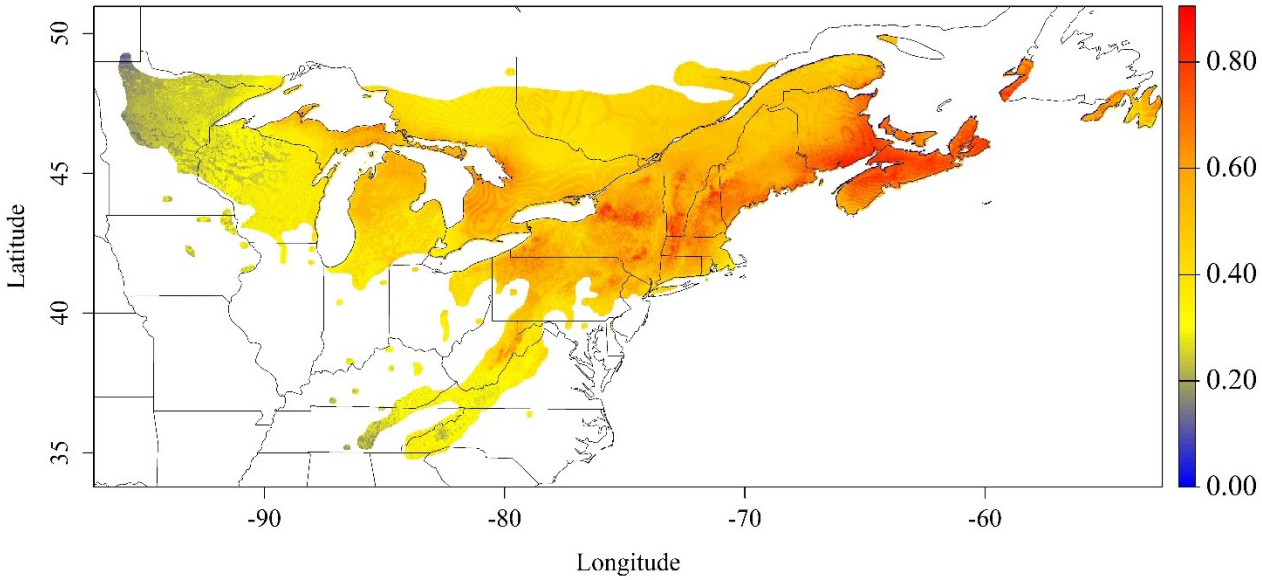


Figure 6. Genetic offset across the range of yellow birch is calculated as the Euclidean distance between the contemporary adaptive landscape (RDA1 + RDA2) and the projected adaptive landscape under the future SSP5-8.5 (2041-2070) climate scenario. Higher values indicate a greater genomic change required to maintain adaptive optima.

Standing genetic variation (SGV) was highest in yellow birch populations from the Western Great Lakes and declined eastward, reaching its lowest levels in the northern Northeastern Appalachian Mountains (Figure 7). In contrast, the population adaptive index (PAI) showed no clear geographic trends. Most populations exhibited moderate PAI values, with scattered outliers displaying lower scores across the range. The highest PAI value was observed in the southern Appalachian Mountain population, which also showed moderate SGV. This same population has been previously identified as an outlier for RDA2-associated allele frequencies.

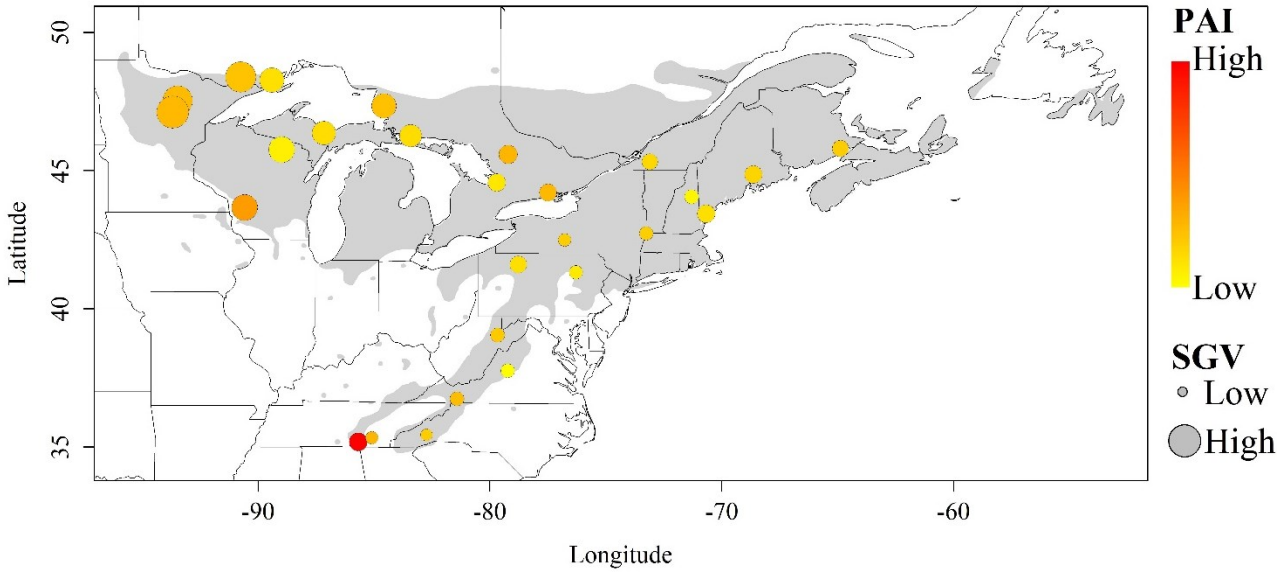
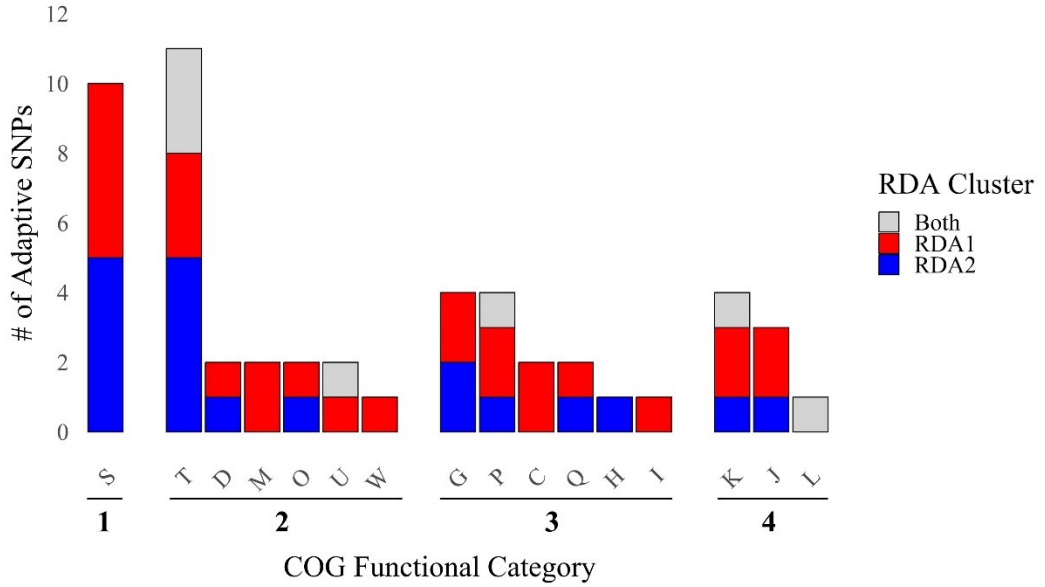


Figure 7. Standing genetic variation (SGV) and population adaptive index (PAI) for the 27 sampled populations. The colour represents the magnitude of the PAI (adaptive deviation from the species-wide mean), while the circle size indicates the level of SGV (within-population genetic diversity at adaptive loci).

Functional categories of putatively adaptive loci

Of the 124 loci identified as putatively adaptive, 87 returned gene function annotations, with 24 associated with RDA1, 18 with RDA2, and 7 associated with both axes. These loci span a wide array of molecular functions and were classified into four broad categories: cellular processes

and signalling, metabolism, information storage and processing, and poorly characterized (Figure 8; see Supplemental Table S2 for detailed gene annotations).



1 POORLY CHARACTERIZED
S: Function unknown

2 CELLULAR PROCESSES & SIGNALING
T: Signal transduction mechanisms
D: Cell cycle control, cell division, chromosome partitioning
M: Cell wall/membrane/envelope biogenesis
O: Posttranslational modification, protein turnover, chaperones
U: Intracellular trafficking, secretion, & vesicular transport
W: Extracellular structures

3 METABOLISM
G: Carbohydrate transport and metabolism
P: Inorganic ion transport & metabolism
C: Energy production and conversion
Q: Secondary metabolites biosynthesis, transport & catabolism
H: Coenzyme transport & metabolism
I: Lipid transport & metabolism

4 INFORMATION STORAGE & PROCESSING
K: Transcription
J: Translation, ribosomal structure & biogenesis
L: Replication, recombination & repair

Figure 8. Distribution of 49 putatively adaptive SNPs across COG functional categories and RDA axes. The stacked barplot shows the number of SNPs under selection under each COG functional category grouped into their respective broad category: (1) Poorly characterized, (2) Cellular processing and signalling, (3) Metabolism, and (4) Information storage and processing. Bars are colour-coded by RDA cluster: RDA-1 associated loci (red), RDA-2 associated loci (blue), and loci associated with both axes (grey). The accompanying legend defined each COG category and its corresponding biological function.

The largest proportion of annotated loci were found to have functions relating to cellular processes and signalling, with signal transduction mechanisms (COG category T) accounting for over 22% of all annotated loci. One such gene, a nodulation receptor kinase-like gene (SymRK) associated with RDA1, plays a well-characterized role in mediating environmental signalling responses (Holsters, 2008; Choudhury & Pandey, 2024). Additionally, four loci encoding serine threonine protein kinases or phosphatases were identified across all three RDA clusters, including the BRI1 gene, which was associated with both axes. Together, they underscore the widespread role of post-translational signalling pathways in regulating climate-associated responses in yellow birch.

Several loci returned functions relating to metabolic processes, including energy production, secondary metabolism, redox regulation, and defence signalling. These loci fell primarily into the functional categories of carbohydrate transport and metabolism (G) and inorganic ion transport and metabolism (P). They include a glycosyltransferase (At5g03795) associated with RDA1 and a geranyl diphosphate synthase (GPPS) associated with RDA2, both of which contribute to secondary metabolite biosynthesis and terpenoid pathways (Hansen *et al.*, 2009; Ali *et al.*, 2020). This group also included genes involved in oxidative stress response, energy balance, and ion homeostasis, including those related to redox enzymes, membrane transporters, and cellular energy exchange. In addition, two loci were annotated with functions potentially involved in pathogen defence and leaf death, suggesting that some metabolic pathways may also contribute to biotic stress resilience in yellow birch.

A small number of loci were associated with information storage and processing, divided between functions relating to transcription (K) and translation and ribosomal biogenesis (J). Within the transcription group, notable loci included a WRKY13 transcription factor associated with

RDA1 and a putative RNA polymerase II C-terminal domain phosphatase-like gene associated with RDA2. These genes may contribute to the regulation of stress-responsive gene expression at both transcriptional and post-transcriptional levels (Zhang *et al*., 2020; Zhang *et al.*, 2022). Among the translation-related loci, one encodes a member of the RNA M5U methyltransferase family involved in ribosomal RNA modification, suggesting a role in translational control under climatic stress.

Several loci returned poorly characterized functional annotations. These included genes with largely unknown functions but potentially relating to stress and hormone signalling, redox and metabolic stress response, and cellular structure, as well as one annotated as a sieve element occlusion (SEOa) gene associated with RDA2 and linked to phloem function (Ernst *et al.*, 2012).

DISCUSSION

Drivers of genetic structure in yellow birch

The RDA variance partitioning revealed that climate explained the largest share of total genetic variance in yellow birch (nearly half of the explainable portion), suggesting that environmental selection has played a central role in shaping population structure. However, the pure climate model lost statistical significance when conditioned on geography and neutral structure, reflecting the well-documented spatial confounding among these predictors (Capblancq *et al.*, 2023). Despite this, the magnitude of variance explained by climate remained substantial, especially when compared to geography and ancestry alone. This mirrors findings in other forest trees, such as red spruce, where climate similarly explained more than either geography or ancestry (Capblancq *et al.*, 2023), but the relatively low confounded variance in yellow birch (~24%) suggests a stronger signal of adaptation, potentially due to its broader range and less strictly north-south distribution. In contrast, confounded variance in red spruce exceeded 45% (Capblancq *et al.*, 2023), and in white pine studies by Nadeau *et al.* (2016), it ranged from 17-40% depending on methodology. This underscores how the ecological breadth of yellow birch may facilitate a cleaner separation of adaptive and demographic signals.

Interestingly, neutral population structure, while explaining the smallest portion of variance (9%), was the only statistically significant pure predictor. This reflects the enduring influence of post-glacial migration and genetic drift, as observed in yellow birch and other temperate forest trees. Distinct haplotype groups have been revealed in yellow birch using cpDNA and linked to multiple southern refugia (Thomson *et al.*, 2015), and similar dynamics have shaped genetic structure in species like *Cercis canadensis* (Ony *et al.*, 2021), *Picea sitchensis* (Mimura & Aitken,

2007), and *Pinus strobus* (Zinck & Rajora, 2016). Notably, no significant substructure was detected at nuclear microsatellites in yellow birch (Thomson *et al.*, 2015), supporting the idea that widespread gene flow, particularly via pollen in wind-pollinated species, buffers against strong geographic differentiation. This pattern is not unique to yellow birch; low genetic differentiation across broad ranges is a near-universal trend in temperate trees (Hamrick *et al.*, 1992), primarily driven by long-distance pollen dispersal and large effective population sizes.

The moderate proportion of shared variance (24%) among climate, geography, and ancestry highlights the spatial autocorrelation inherent in landscape genomic data, where adaptation, historical processes, and spatial structure are deeply intertwined. This collinearity is a known challenge in landscape genomics and is especially pronounced in species that recolonized from southern refugia after the last glacial maximum (Hewitt, 1999; Jones *et al.*, 2013). In such contexts, controlling for population structure may reduce false positives, but it can also obscure real signals of adaptation, particularly when climate and ancestry covary (Capblancq *et al.*, 2023). As Nadeau *et al.* (2016) noted, even advanced GEA methods struggle to partition these effects cleanly. Together, these results suggest that while adaptive differentiation driven by climate is strong in yellow birch, it remains partly entangled with geography and ancestry, as is typical of post-glacial temperate forest species (Hewitt 199; Aitken *et al.*, 2008; Sork *et al.*, 2010; Capblancq *et al.*, 2023).

Climate associations of adaptive loci

The adaptively enriched RDA identified two primary climate gradients associated with genetic variation in yellow birch. The first axis corresponded to extreme minimum temperature, degree-days below 0°C, and winter precipitation, collectively representing a gradient of winter

extremeness and duration. The second axis was primarily associated with precipitation as snow, capturing variation in snowpack conditions across the range. These axes reflect distinct components of winter stress, suggesting that selection on traits related to cold tolerance, dormancy regulation, and snow-mediated insulation may underlie observed patterns of genomic differentiation. The clustering of adaptive loci along these axes mirrors patterns observed in other temperate species, including *Picea rubens* (Capblancq *et al*., 2023), *Pinus strobus* (Li *et al.*, 1997), and *Pinus contorta* (Mahony *et al.*, 2020), where allele frequencies tracked environmental variation in cold exposure and frost-related injury. Similarly, in *Pinus sylvestris,* genotype-environment associations have been shaped by winter severity and growing-season temperature, underscoring how climatic extremes across broad latitudinal gradients can drive adaptive divergence (Calleja-Rodriguez *et al.*, 2019).

In addition to the two main clusters of loci associated with the separate axes, a third set of nine loci was found to be strongly associated with both climate gradients. This overlap suggests that certain genomic regions are influenced by multi-dimensional selection, responding simultaneously to variation in both winter temperature severity and snowfall. Alternatively, it is possible that some loci have pleiotropic effects, contributing to multiple climate-related traits simultaneously.

The climate variables identified in this study, particularly extreme minimum temperature and snowfall, highlight adaptation to overwinter environmental conditions. This complements earlier provenance and trait-based studies in yellow birch that identified clinal patterns of local adaptation across geographic and climatic gradients. For example, Clausen (1968b) observed latitudinal variation in growth cessation correlated with growing season length and temperature, while Wearstler and Barner (1977) reported north-south clines in height and germination traits

across provenances. More recently, Maloney *et al.*, (2024) demonstrated that leaf morphology and water use efficiency varied predictably with growing-season temperature, precipitation, and heat-moisture balance. Taken together, these studies suggest that adaptation in yellow birch operates along multiple climatic axes. Whereas earlier studies emphasize responses to growing-season conditions, the genomic findings here point to winter climate as a key selective driver. Together, these findings suggest that different seasonal climate factors influence distinct components of adaptation in yellow birch, with functional traits responding to moisture stress during the growing season and genomic variation reflecting selection pressures associated with winter severity.

Spatial patterns of adaptation and future vulnerability

The spatial distribution of adaptive genetic variation in yellow birch revealed distinct regional patterns associated with winter-related climate variables. The RDA-1 based adaptive index, which reflects adaptation to winter severity, duration and precipitation, displayed a strong latitudinal gradient, with higher adaptation to longer, colder and dryer winters in the northern portion of the range. This suggests that populations in the boreal-northern zones have been shaped by selection for traits promoting cold and frost resilience. In contrast, the RDA-2 based index, linked to adaptation to high snowfall, exhibited both latitudinal and longitudinal structure, with higher adaptation to high snowfall in the Maritimes, Laurentians and high-elevation regions of the northeastern United States. Although these regions receive substantial snowfall, snowpack may be less persistent to due frequent mid-winter thaws, costal climatic influences, or rapid spring melt, leading to increased exposure to freeze-thaw cycles. As such, populations in these areas may be subject to stronger selection for traits that confer tolerance to ephemeral or reduced snow cover, such as increased freezing resistance or the ability to withstand frequent freeze-thaw transitions.

Alternatively, given yellow birch's preference for moist microsites, late-season snowmelt and associated early-summer moisture availability may exert selection for traits that enhance early growth and water use efficiency in these high-snowfall environments. Similar patterns of climate-associated adaptation have been observed in other temperate hardwoods, including *Acer saccharum*, which exhibits reduced growth in areas with low snowpack (Reinmann *et al.*, 2019).

Future climate projections under a high-emission scenario (2041-2070) indicate that the adaptive landscape of yellow birch is likely to shift, particularly along axes related to winter severity and snowfall. Projected increases in RDA-1 index values suggest an amplification of differentiation in winter adaptation across the range, with notable shifts in the Northeastern Appalachians, the St. Lawrence valley, and the central and eastern Great Lakes regions. For RDA-2, projected changes indicate a growing mismatch between present-day snowfall-associated adaptation and future conditions in the Maritimes, Northeastern Appalachians, southern Ontario, and Michigan. These regions may face an increased risk of maladaptation, particularly among populations currently specialized for historical snowfall regimes. Comparable patterns of climate-induced vulnerability have been reported in species such as *Quercus robur* (Leroy *et al.*, 2019) and *Pinus sylvestris* (Hallingbäck *et al.*, 2021), where population fitness declined as environmental conditions diverged from historical selection pressures. In yellow birch, such mismatches could lead to declining adaptive fit, especially if local standing genetic variation is insufficient to accommodate rapid environmental shifts.

These results reinforce broader patterns observed in temperate hardwoods, where genomic differentiation is often tightly coupled to regional climate conditions (Keller *et al.*, 2012; Gugger *et al.*, 2021; Capblancq *et al.*, 2023). In yellow birch, cold- and snow-adapted populations, especially those in northern latitudes or higher elevations, may represent vulnerable edges and

valuable genetic reservoirs under future climates. Prioritizing these populations for monitoring, *in situ* protection, or potential assisted migration could help maintain adaptive capacity and resilience within the species. As climate change continues to reshape ecological and evolutionary landscapes, conservation strategies rooted in genomic data will be essential for safeguarding adaptive potential and long-term persistence in this foundational forest species.

Genetic offset & adaptive capacity

Patterns of genetic offset in yellow birch varied geographically, with the highest projected shifts in adaptive allele frequencies occurring in the Maritimes and the Northeastern Appalachian Mountains, and the lowest offset values observed in the western Great Lakes and southern Appalachians. While widespread genomic change may be required to track future climates, some populations may already harbour alleles pre-aligned with future conditions. The overall modest offset across much of the range suggests that yellow birch may possess broad adaptive potential, possibly buffered by its allohexaploid genome, which enhances genomic plasticity and may reduce the magnitude of change required to maintain local adaptation. Similar buffering effects have been observed in other polyploid plant systems, enhancing responsiveness to environmental stress by facilitating the up- or down-regulation of gene expression (Chen, 2007; Jackson & Chen, 2010); and in *Arabidopsis thaliana*, polyploids have shown greater phenotypic plasticity than their diploid progenitors, enabling more flexible responses to climate variation (Mattingly & Hovick, 2023).

In the western Great Lakes, where offset values were particularly low, historic and ongoing introgression from *Betula papyrifera* (Thomson, 2015) and possibly *B. pumila* (Danick & Barnes, 1975b; Barnes & Danick, 1985) may have contributed to the presence of alleles aligned with future

climatic conditions. Additionally, ongoing hybridization in these regions may enhance the potential for rapid evolutionary responses by increasing standing genetic variation and facilitating the spread of adaptive alleles. While such processes do not directly reduce the magnitude of predicted offset, they may help populations track environmental change more effectively over time. This gene flow may be further facilitated by hexaploidy, which can relax reproductive barriers and enhance genomic integration (Schmickl & Yant, 2021). Similar dynamics have been reported in *Quercus petraea*, where gene flow from *Q. robur* introduced climate-adaptive alleles that contributed to clinal genomic variation (Leroy *et al.*, 2019), and in an Icelandic birch, where introgression between *B. nana* and *B. pubescens* supported resilience to climate warming (Anamthawat-Jónsson, 2019). However, hybridization does not always result in adaptive benefits; crosses between *B. alleghaniensis* and *B. lenta* have shown reduced vigor and germination, suggesting postzygotic barriers can constrain certain hybrid combinations (Sharik & Barnes, 1971).

Measures of standing genetic variation (SGV) and the population adaptive index (PAI) offer complementary insights into yellow birch's adaptive capacity. SGV, which reflects the existing genetic diversity at climate-associated loci, was highest in the western Great Lakes and declined eastward, consistent with the regions of past introgression (Thomson *et al*., 2015). In contrast, PAI, which measures how divergent a population's adaptive allele frequencies are from the mean, showed no consistent spatial patterns, likely reflecting the influence of localized selection pressure or recent demographic history. Notably, a southern Appalachian outlier exhibited the highest PAI and relatively high SGV, despite its position at the warm, low-snowfall edge of the range. This unexpected combination may reflect microclimatic effects or a unique

historical legacy and stands in contrast to other nearby southern edge populations with lower SGV and moderate PAI.

Different spatial patterns of variation in SGV and PAI highlight how these metrics capture different facets of adaptive potential, where SGV represents the raw material for response, while PAI reflects realized adaptive divergence (Capblancq *et al.*, 2020). Evaluating both metrics together helps identify populations that are not only distinct but also poised or poorly positioned to respond to future environmental changes. These results suggest that while some populations, particularly in the western Great Lakes, may retain a strong capacity to respond to future change, others, like populations in the Northern Appalachians may be more constrained by low genetic variation or past selective filtering, emphasizing the uneven distribution of adaptive potential across the range.

Gene functions & local adaptation

Putative adaptive loci in yellow birch span a diverse array of biological functions, reflecting the multifaceted nature of climate adaptation. Function categories included signal transduction, hormonal regulation, secondary metabolism, transcriptional control and RNA processing, indicating that adaptation involves coordinated responses across multiple physiological systems. This diversity mirrors findings in other temperate trees where adaptation to climate gradients involves both regulatory and structural genes acting across multiple biological pathways (Capblancq *et al.*, 2023; Meger *et al.*, 2024). Several loci were linked to known stress-related gene families, including transcription factors (WRKY13; Chen *et al.*, 2012), receptor kinases (SymRK, BRI1; Ye *et al.*, 2017), heat shock proteins (HSP70; Zhang *et al.*, 2025), redox

and energy metabolism (GPPS, At5g03795; Rehman *et al.*, 2022; Sinha *et al.*, 2024), and post-transcriptional regulators such as RNA methyltransferases (Cai *et al.*, 2025) and pentatricopeptide repeat (PPR) proteins (Xing *et al.*, 2018). Together, this breadth of function suggests that climate adaptation in yellow birch engages multiple coordinated mechanisms to manage temperature and moisture stress.

Within this diverse gene pool, several key functional groups stand out as likely contributors to climatic resilience. Transcription factors, including WRKY13, play central roles in mediating abiotic and biotic stress responses and have been shown to be upregulated or functionally involved in drought, salt, cold, and pathogen stress in multiple species, including *Oryza sativa* (Xiao *et al.*, 2013) and *Paeonia lactiflora* (Wang *et al.*, 2019). Several loci encoded serine/threonine receptor-like kinases/phosphatases, key components of stress-responsive hormonal pathways such as brassinosteroid and ABA signalling (Khan *et al.*, 2023). One of these, BRI1, a canonical brassinosteroid receptor, exemplifies this group's role in abiotic stress signalling and developmental regulation, with demonstrated involvement in xylem differentiation and adaptation to drought and temperature fluctuations in *Quercus* (Ai *et al.*, 2023) and *Picea abies* (Wang *et al.*, 2021). SymRK-like receptor kinases, best known for their essential role in legume-rhizobia symbiosis, also modulate G-protein signalling pathways via phosphorylations on RGS proteins, contributing to the fine-tuning of downstream responses that are increasingly recognized as intersecting with broader abiotic and biotic stress signalling networks in plants (Holsters, 2008; Choudhury & Pandey, 2024). The presence of GPPS and At5g03795 (a glycosyltransferase) among adaptive loci suggests a role for secondary metabolism in stress resilience, GPPS has been shown to contribute to terpenoid biosynthesis under abiotic stress in cotton (Ali *et al.*, 2020), and At5g03795 has been linked to cell wall modification (Hansen *et al*., 2009). Meanwhile, regulatory

genes such as RNA polymerase II and RNA M5U methyltransferases highlight the importance of post-transcriptional and epitranscriptomic mechanisms in stress adaptation, with roles in modulating gene expression under salt and heat stress (Zhang *et al.*, 2020; Zhang *et al.*, 2022).

The clustering of gene functions by RDA axes suggests that different climate gradients may select for distinct molecular strategies. Genes associated with winter temperature extremes, duration and winter precipitation included WRKY13 and At5g03795, while genes aligned with snowfall variation featured genes like GPPS and RNA regulatory elements. This separation implied that climate adaptation in yellow birch is both axis-specific and modular, with different environmental pressures targeting different aspects of the genomic response. The identification of genes involved in environmental sensing, signal integration, metabolic buffering, and gene regulation underscores that adaptation in yellow birch is polygenic and multidimensional, not the result of selection on a single pathway. These findings point to promising targets for future validation through expression profiling, functional genomics, or landscape transcriptomics and highlight the value of incorporating gene-level insights into broader models of climate resilience in temperate forest trees.

CONCLUSION

Local adaptation in yellow birch reflects a complex interplay of demographic history and climate-driven selection across its range. Structured genetic variation aligned with geographic and climate gradients points to the combined influence of post-glacial processes and spatially varying selection. Climate emerged as a dominant driver of adaptive genetic structure, with clear associations between allele frequency shifts and winter-related stress gradients. The identification of diverse candidate loci, including genes involved in transcriptional regulation, signal transduction, secondary metabolism and RNA processing, highlights the polygenic and multilayered nature of local adaptation in this temperate hardwood. By linking functional genetic variation to biologically meaningful climate axes, this study provides a valuable foundation for climate-informed seed sourcing, assisted migration strategies, and adaptive forest management. More broadly, it underscores the importance of integrating landscape genomics and functional annotation to anticipate how forest tree species may respond to accelerating environmental change.

ADDITIONAL INFORMATION

AI disclosure

OpenAI ChatGPT (GPT-4, 2025) was used to assist with code development, data processing, figure customization, and language editing during the preparation of the manuscript. All outputs were critically reviewed, modified as needed, and fully approved by the lead author. No content was generated autonomously or used without verification, and all analytical and interpretive decisions were made by the author(s).

Data availability

All code used for data processing and analysis has been made publicly available via Dryad. This includes R scripts for polyRAD filtering, partitioning of variance, GEA analyses, and index projection. VCF files used in the analyses are also included in the repository. The dataset can be accessed at [To be inserted: Dryad link] (DOI: To be inserted).

LITERATURE CITED

Ai, W., Liu, H., Wang, Yutao, Wang, Yu, Wei, J., Zhang, X., Lu, X., 2023. Identification of Functional Brassinosteroid Receptor Genes in Oaks and Functional Analysis of QmBRI1. IJMS 24, 16405. https://doi.org/10.3390/ijms242216405

Aitken, S.N., Whitlock, M.C., 2013. Assisted Gene Flow to Facilitate Local Adaptation to Climate Change. Annu. Rev. Ecol. Evol. Syst. 44, 367–388. https://doi.org/10.1146/annurev-ecolsys-110512-135747

Aitken, S.N., Yeaman, S., Holliday, J.A., Wang, T., Curtis-McLane, S., 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations: Climate change outcomes for tree populations. Evolutionary Applications 1, 95–111. https://doi.org/10.1111/j.1752-4571.2007.00013.x

Ali, F., Qanmber, G., Wei, Z., Yu, D., Li, Y.H., Gan, L., Li, F., Wang, Z., 2020. Genome-wide characterization and expression analysis of geranylgeranyl diphosphate synthase genes in cotton (Gossypium spp.) in plant development and abiotic stresses. BMC Genomics 21, 561. https://doi.org/10.1186/s12864-020-06970-8

Allen, C.D., Macalady, A.K., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M., Kitzberger, T., Rigling, A., Breshears, D.D., Hogg, E.H. (Ted), Gonzalez, P., Fensham, R., Zhang, Z., Castro, J., Demidova, N., Lim, J.-H., Allard, G., Running, S.W., Semerci, A., Cobb, N., 2010. A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. Forest Ecology and Management 259, 660–684. https://doi.org/10.1016/j.foreco.2009.09.001

Anamthawat-Jónsson, K., 2019. Hybrid introgression: the outcomes of gene flow in birch. ScienceAsia 45, 203. https://doi.org/10.2306/scienceasia1513-1874.2019.45.203

Ashburner, K., McAllister, H.A., Rix, M., 2013. The genus Betula: a taxonomic revision of birches, A botanical magazine monograph. Kew publ, Kew.

Barnes, B.V., Dancik, B.P., 1985. Characteristics and origin of a new birch species, Betula murrayana , from southeastern Michigan. Can. J. Bot. 63, 223–226. https://doi.org/10.1139/b85-025

Bonin, A., Nicole, F., Pompanon, F., Miaud, C., Taberlet, P., 2007. Population Adaptive Index: a New Method to Help Measure Intraspecific Genetic Diversity and Prioritize Populations for Conservation. Conservation Biology 21, 697–708. https://doi.org/10.1111/j.1523-1739.2007.00685.x

Cai, J., Shen, L., Kang, H., Xu, T., 2025. RNA modifications in plant adaptation to abiotic stresses. Plant Communications 6, 101229. https://doi.org/10.1016/j.xplc.2024.101229

Calleja-Rodriguez, A., Andersson Gull, B., Wu, H.X., Mullin, T.J., Persson, T., 2019. Genotype-by-environment interactions and the dynamic relationship between tree vitality and height in northern Pinus sylvestris. Tree Genetics & Genomes 15, 36. https://doi.org/10.1007/s11295-019-1343-8

Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., Huerta-Cepas, J., 2021. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Molecular Biology and Evolution 38, 5825–5829. https://doi.org/10.1093/molbev/msab293

Capblancq, T., Forester, B.R., 2021. Redundancy analysis: A Swiss Army Knife for landscape genomics. Methods Ecol Evol 12, 2298–2309. https://doi.org/10.1111/2041-210X.13722

Capblancq, T., Lachmuth, S., Fitzpatrick, M.C., Keller, S.R., 2023. From common gardens to candidate genes: exploring local adaptation to climate in red spruce. New Phytologist 237, 1590–1605. https://doi.org/10.1111/nph.18465

Capblancq, T., Morin, X., Gueguen, M., Renaud, J., Lobreaux, S., Bazin, E., 2020. Climate-associated genetic variation in Fagus sylvatica and potential responses to climate change in the French Alps. J of Evolutionary Biology 33, 783–796. https://doi.org/10.1111/jeb.13610

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology 22, 3124–3140. https://doi.org/10.1111/mec.12354

Chen, L., Song, Y., Li, S., Zhang, L., Zou, C., Yu, D., 2012. The role of WRKY transcription factors in plant abiotic stresses. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1819, 120–128. https://doi.org/10.1016/j.bbagrm.2011.09.002

Chen, Z.J., 2007. Genetic and Epigenetic Mechanisms for Gene Expression and Phenotypic Variation in Plant Polyploids. Annu. Rev. Plant Biol. 58, 377–406. https://doi.org/10.1146/annurev.arplant.58.032806.103835

Chhatre, V.E., Fetter, K.C., Gougherty, A.V., Fitzpatrick, M.C., Soolanayakanahally, R.Y., Zalesny, R.S., Keller, S.R., 2019. Climatic niche predicts the landscape structure of locally adaptive standing genetic variation. https://doi.org/10.1101/817411

Choudhury, S.R., Pandey, S., 2024. SymRK Regulates G-Protein Signaling During Nodulation in Soybean (Glycine max) by Modifying RGS Phosphorylation and Activity. MPMI 37, 765–775. https://doi.org/10.1094/MPMI-04-24-0036-R

Clark, L.V., Lipka, A.E., Sacks, E.J., 2019. polyRAD: Genotype Calling with Uncertainty from Sequencing Data in Polyploids and Diploids. G3 Genes|Genomes|Genetics 9, 663–673. https://doi.org/10.1534/g3.118.200913

Clark, L.V., Mays, W., Lipka, A.E., Sacks, E.J., 2022. A population-level statistic for assessing Mendelian behavior of genotyping-by-sequencing data from highly duplicated genomes. BMC Bioinformatics 23, 101. https://doi.org/10.1186/s12859-022-04635-9

Clausen, K.E., 1980. Survival, growth, and flowering of yellow birch progenies in an open-field test. Silvae Genet 29, 108–114.

Clausen, K.E., 1975. Variation in early growth and survival of yellow birch provenances. In Tree improvement and genetics——Northeastern Forest Tree Improvement Conference——1974. Syracuse, NY 139–148.

Clausen, K.E., 1968a. Natural variation in catkin and fruit characteristics of yellow birch. In Proceedings of the 15th Northeastern Forest Tree Improvement Conference. Morgantown, West Virginia. 2–7.

Clausen, K.E., 1968b. Variation in height growth and growth cessation of 55 yellow birch seed sources. In 8th Lake States Forest Tree Improvement Conference. North Central Forest Experiment Station, USDA Forest Service, Madison, WI 0–3.

Dancik, B.P. and Barnes, B.V., 1975b. Multivariate analyses of hybrid populations. Naturaliste Canadien V., 19760631568, 835–843.

Dancik, B.P., Barnes, B.V., 1975a. Leaf Variability in Yellow Birch (Betula alleghaniensis) in Relation to Environment. Can. J. For. Res. 5, 149–159. https://doi.org/10.1139/x75-021

De Ronne, M., Légaré, G., Belzile, F., Boyle, B., Torkamaneh, D., 2023. 3D-GBS: a universal genotyping-by-sequencing approach for genomic selection and other high-throughput low-cost applications in species with small to medium-sized genomes. Plant Methods 19, 13. https://doi.org/10.1186/s13007-023-00990-7

Ellis, N., Smith, S.J., Pitcher, C.R., 2012. Gradient forests: calculating importance gradients on physical predictors. Ecology 93, 156–168. https://doi.org/10.1890/11-0252.1

Erdmann, G.G., 1990. Betula alleghaniensis Britton. Yellow birch., in: Silvics of North America. Wolume 2 Hardwoods. USDA Forest Service Agricultural Handbook. Washington, D.C., pp. 133–147.

Eric Normandeau, n.d. STACKS Workflow2: RADseq workflow using STACKS2 GitHub.

Ernst, A.M., Jekat, S.B., Zielonka, S., Müller, B., Neumann, U., Rüping, B., Twyman, R.M., Krzyzanek, V., Prüfer, D., Noll, G.A., 2012. Sieve element occlusion (SEO) genes encode structural phloem proteins involved in wound sealing of the phloem. Proc. Natl. Acad. Sci. U.S.A. 109. https://doi.org/10.1073/pnas.1202999109

Fitzpatrick, M.C., Keller, S.R., 2015. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. Ecology Letters 18, 1–16. https://doi.org/10.1111/ele.12376

Forester, B.R., Jones, M.R., Joost, S., Landguth, E.L., Lasky, J.R., 2016. Detecting spatial genetic signatures of local adaptationin heterogeneous landscapes. Molecular Ecology 25, 104–120.

Forester, B.R., Lasky, J.R., Wagner, H.H., Urban, D.L., 2018. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. Molecular Ecology 27, 2215–2233. https://doi.org/10.1111/mec.14584

Gugger, P.F., Fitz-Gibbon, S.T., Albarrán-Lara, A., Wright, J.W., Sork, V.L., 2021. Landscape genomics of Quercus lobata reveals genes involved in local climate adaptation at multiple spatial scales. Molecular Ecology 30, 406–423. https://doi.org/10.1111/mec.15731

Günther, T., Coop, G., 2013. Robust Identification of Local Adaptation from Allele Frequencies. Genetics 195, 205–220. https://doi.org/10.1534/genetics.113.152462

Hallingbäck, H.R., Burton, V., Vizcaíno-Palomar, N., Trotter, F., Liziniewicz, M., Marchi, M., Berlin, M., Ray, D., Benito Garzón, M., 2021. Managing Uncertainty in Scots Pine Range-Wide Adaptation Under Climate Change. Front. Ecol. Evol. 9, 724051. https://doi.org/10.3389/fevo.2021.724051

Hamrick, J.L., Godt, M.J.W., Sherman-Broyles, S.L., 1992. Factors influencing levels of genetic diversity in woody plant species, in: Adams, W.T., Strauss, S.H., Copes, D.L., Griffin, A.R. (Eds.), Population Genetics of Forest Trees, Forestry Sciences. Springer Netherlands, Dordrecht, pp. 95–124. https://doi.org/10.1007/978-94-011-2815-5_7

Hansen, S.F., Bettler, E., Wimmerová, M., Imberty, A., Lerouxel, O., Breton, C., 2009. Combination of Several Bioinformatics Approaches for the Identification of New Putative Glycosyltransferases in Arabidopsis. J. Proteome Res. 8, 743–753. https://doi.org/10.1021/pr800808m

Hewitt, G.M., 1999. Post-glacial re-colonization of European biota. Biological Journal of the Linnean Society 68, 87–112. https://doi.org/10.1111/j.1095-8312.1999.tb01160.x

Holliday, J.A., Aitken, S.N., Cooke, J.E.K., Fady, B., González-Martínez, S.C., Heuertz, M., Jaramillo-Correa, J.-P., Lexer, C., Staton, M., Whetten, R.W., Plomion, C., 2017. Advances in ecological genomics in forest trees and applications to genetic resources conservation and breeding. Mol Ecol 26, 706–717. https://doi.org/10.1111/mec.13963

Holmes, R.T., Schultz, J.C., 1988. Food availability for forest birds: effects of prey distribution and abundance on bird foraging. Can. J. Zool. 66, 720–728. https://doi.org/10.1139/z88-107

Holsters, M., 2008. SYMRK, an enigmatic receptor guarding and guiding microbial endosymbioses with plant roots. Proc. Natl. Acad. Sci. U.S.A. 105, 4537–4538. https://doi.org/10.1073/pnas.0801270105

Isabel, N., Holliday, J.A., Aitken, S.N., 2020. Forest genomics: Advancing climate adaptation, forest health, productivity, and conservation. Evol Appl 13, 3–10. https://doi.org/10.1111/eva.12902

Jackson, S., Chen, Z.J., 2010. Genomic and expression plasticity of polyploidy. Current Opinion in Plant Biology 13, 153–159. https://doi.org/10.1016/j.pbi.2009.11.004

Jones, M.R., Forester, B.R., Teufel, A.I., Adams, R.V., Anstett, D.N., Goodrich, B.A., Landguth, E.L., Joost, S., Manel, S., 2013. INTEGRATING LANDSCAPE GENOMICS AND SPATIALLY EXPLICIT APPROACHES TO DETECT LOCI UNDER SELECTION IN CLINAL POPULATIONS: SPECIAL SECTION. Evolution 67, 3455–3468. https://doi.org/10.1111/evo.12237

Keller, S.R., Levsen, N., Olson, M.S., Tiffin, P., 2012. Local Adaptation in the Flowering-Time Gene Network of Balsam Poplar, Populus balsamifera L. Molecular Biology and Evolution 29, 3143–3152. https://doi.org/10.1093/molbev/mss121

Khan, T.A., Kappachery, S., Karumannil, S., AlHosani, M., Almansoori, N., Almansoori, H., Yusuf, M., Tran, L.-S.P., Gururani, M.A., 2023. Brassinosteroid Signaling Pathways: Insights into Plant Responses under Abiotic Stress. IJMS 24, 17246. https://doi.org/10.3390/ijms242417246

Lachmuth, S., Capblancq, T., Prakash, A., Keller, S.R., Fitzpatrick, M.C., 2023. Novel genomic offset metrics account for local adaptation in climate suitability forecasts and inform assisted migration. https://doi.org/10.1101/2023.06.05.541958

Láruson, Á.J., Fitzpatrick, M.C., Keller, S.R., Haller, B.C., Lotterhos, K.E., 2022. Seeing the forest for the trees: Assessing genetic offset predictions from gradient forest. Evol Appl 15, 403–416. https://doi.org/10.1111/eva.13354

Lavania, U.C., 2020. Plant speciation and polyploidy: in habitat divergence and environmental perspective. Nucleus 63, 1–5. https://doi.org/10.1007/s13237-020-00311-6

Leites, L.P., Rehfeldt, G.E., Steiner, K.C., 2019. Adaptation to climate in five eastern North America broadleaf deciduous species: Growth clines and evidence of the growth-cold tolerance trade-off. Perspectives in Plant Ecology, Evolution and Systematics 37, 64–72. https://doi.org/10.1016/j.ppees.2019.02.002

Leroy, T., Louvet, J.-M., Lalanne, C., Le Provost, G., Labadie, K., Aury, J.-M., Delzon, S., Plomion, C., Kremer, A., 2019. Adaptive introgression as a driver of local adaptation to climate in European white oaks. https://doi.org/10.1101/584847

Li, P., Beaulieu, J., Daoust, G., Plourde, A., 1997. Patterns of adaptive genetic variation in eastern white pine (Pinus strobus) from Quebec. Can. J. For. Res. 27, 199–206. https://doi.org/10.1139/x96-158

Liang, J., Crowther, T.W., Picard, N., Wiser, S., Zhou, M., Alberti, G., Schulze, E.-D., McGuire, A.D., Bozzato, F., Pretzsch, H., de-Miguel, S., Paquette, A., Hérault, B., Scherer-Lorenzen, M., Barrett, C.B., Glick, H.B., Hengeveld, G.M., Nabuurs, G.-J., Pfautsch, S., Viana, H., Vibrans, A.C., Ammer, C., Schall, P., Verbyla, D., Tchebakova, N., Fischer, M., Watson, J.V., Chen, H.Y.H., Lei, X., Schelhaas, M.-J., Lu, H., Gianelle, D., Parfenova, E.I., Salas, C., Lee, E., Lee, B., Kim, H.S., Bruelheide, H., Coomes, D.A., Piotto, D., Sunderland, T., Schmid, B., Gourlet-Fleury, S., Sonké, B., Tavani, R., Zhu, J., Brandl, S., Vayreda, J., Kitahara, F., Searle, E.B., Neldner, V.J., Ngugi, M.R., Baraloto, C., Frizzera, L., Bałazy, R., Oleksyn, J., Zawiła-Niedźwiecki, T., Bouriaud, O., Bussotti, F., Finér, L., Jaroszewicz, B., Jucker, T., Valladares, F., Jagodzinski, A.M., Peri, P.L., Gonmadje, C., Marthy, W., O'Brien, T., Martin, E.H., Marshall, A.R., Rovero, F., Bitariho, R., Niklaus, P.A., Alvarez-Loayza, P., Chamuya, N., Valencia, R., Mortier, F., Wortel, V., Engone-Obiang, N.L., Ferreira, L.V., Odeke, D.E., Vasquez, R.M.,

Lewis, S.L., Reich, P.B., 2016. Positive biodiversity-productivity relationship predominant in global forests. Science 354, aaf8957. https://doi.org/10.1126/science.aaf8957

Long, S., Jiang, J., Lin, H., 2025. Whole-Genome Identification and Expression Profiles of WRKY Genes Related to the Leaf Expansion Period in the Camphor Tree. Forests 16, 266. https://doi.org/10.3390/f16020266

Lotterhos, K.E., Whitlock, M.C., 2015. The relative power of genome scans to detect localadaptation depends on sampling design and statisticalmethod. Molecular Ecology 24, 1031–1046.

Mahony, C.R., MacLachlan, I.R., Lind, B.M., Yoder, J.B., Wang, T., Aitken, S.N., 2020. Evaluating genomic data for management of local adaptation in a changing climate: A lodgepole pine case study. Evolutionary Applications 13, 116–131. https://doi.org/10.1111/eva.12871

Mahony, C.R., Wang, T., Hamann, A., Cannon, A.J., 2022. A global climate model ensemble for downscaled monthly climate normals over North America. Intl Journal of Climatology 42, 5871–5891. https://doi.org/10.1002/joc.7566

Maloney, A., Dang, Q.L., Godakanda, P.M., Thomson, A., 2024. Genetic variation in growth and leaf traits associated with local adaptation to climate in yellow birch (Betula alleghaniensis Britton). Botany 102, 198–210. https://doi.org/10.1139/cjb-2023-0095

Maloney, A.S., 2022. Genetic variation of functional traits related to drought tolerance in yellow birch seedlings (thesis). Knowledge Commons - Lakehead University 126.

Meek, M.H., Beever, E.A., Barbosa, S., Fitzpatrick, S.W., Fletcher, N.K., Mittan-Moreau, C.S., Reid, B.N., Campbell-Staton, S.C., Green, N.F., Hellmann, J.J., 2023. Understanding Local Adaptation to Prepare Populations for Climate Change. BioScience 73, 36–47. https://doi.org/10.1093/biosci/biac101

Meger, J., Ulaszewski, B., Chmura, D.J., Burczyk, J., 2024. Signatures of local adaptation to current and future climate in phenology-related genes in natural populations of Quercus robur. BMC Genomics 25, 78. https://doi.org/10.1186/s12864-023-09897-y

Menzel, J.M., Ford, W.M., Edwards, J.W., Menzel, M.A., 2004. Nest Tree Use by the Endangered Virginia Northern Flying Squirrel in the Central Appalachian Mountains. The American Midland Naturalist 151, 355–368. https://doi.org/10.1674/0003-0031(2004)151[0355:NTUBTE]2.0.CO;2

Mimura, M., Aitken, S.N., 2007. Adaptive gradients and isolation-by-distance with postglacial migration in Picea sitchensis. Heredity 99, 224–232. https://doi.org/10.1038/sj.hdy.6800987

Muniz, A.C., De Lemos-Filho, J.P., Lovato, M.B., 2024. Non-adaptedness and vulnerability to climate change threaten Plathymenia trees (Fabaceae) from the Cerrado and Atlantic Forest. Sci Rep 14, 25611. https://doi.org/10.1038/s41598-024-75664-y

Nadeau, S., Meirmans, P.G., Aitken, S.N., Ritland, K., Isabel, N., 2016. The challenge of separating signatures of local adaptation from those of isolation by distance and colonization history: The case of two white pines. Ecology and Evolution 6, 8649–8664. https://doi.org/10.1002/ece3.2550

Ony, M., Klingeman, W.E., Zobel, J., Trigiano, R.N., Ginzel, M., Nowicki, M., Boggess, S.L., Everhart, S., Hadziabdic, D., 2021. Genetic diversity in North American Cercis Canadensis reveals an ancient population bottleneck that originated after the last glacial maximum. Sci Rep 11, 21803. https://doi.org/10.1038/s41598-021-01020-z

Park, A., Talbot, C., Smith, R., 2018. Trees for tomorrow: an evaluation framework to assess potential candidates for assisted migration to Manitoba's forests. Climatic Change 148, 591–606. https://doi.org/10.1007/s10584-018-2201-7

Pedlar, J.H., McKenney, D.W., Lu, P., 2021. Critical seed transfer distances for selected tree species in eastern North America. Journal of Ecology 109, 2271–2283. https://doi.org/10.1111/1365-2745.13605

Plummer, M., 2024. rjags: Bayesian Graphical Models using MCMC. R package version 4-17.
Plummer, M., 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing. Vienna, Austria.
Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033
Rehman, H.M., Khan, U.M., Nawaz, S., Saleem, F., Ahmed, N., Rana, I.A., Atif, R.M., Shaheen, N., Seo, H., 2022. Genome Wide Analysis of Family-1 UDP Glycosyltransferases in Populus trichocarpa Specifies Abiotic Stress Responsive Glycosylation Mechanisms. Genes 13, 1640. https://doi.org/10.3390/genes13091640
Reinmann, A.B., Susser, J.R., Demaria, E.M.C., Templer, P.H., 2019. Declines in northern forest tree growth following snowpack decline and soil freezing. Global Change Biology 25, 420–430. https://doi.org/10.1111/gcb.14420
Rellstab, C., Zoller, S., Walthert, L., Lesur, I., Pluess, A.R., Graf, R., Bodénès, C., Sperisen, C., Kremer, A., Gugerli, F., 2016. Signatures of local adaptation in candidate genes of oaks (Quercus spp.) with respect to present and future climatic conditions. Mol Ecol 25, 5907–5924. https://doi.org/10.1111/mec.13889
Rice, A., Šmarda, P., Novosolov, M., Drori, M., Glick, L., Sabath, N., Meiri, S., Belmaker, J., Mayrose, I., 2019. The global biogeography of polyploid plants. Nat Ecol Evol 3, 265–273. https://doi.org/10.1038/s41559-018-0787-9
Sang, Y., Long, Z., Dan, X., Feng, J., Tingting, Shi, Jia, Changfu, Zhang, Xinxin, Lai, Qiang, Yang, Guanglei, Zhang, Hongying, Xu, Xiaoting, Liu, Huanhuan, Jiang, Yuanzhong, Ingvarsson, Par K., Liu, Jianquan, Mao, Kangshan, Wang, Jing, 2022. Genomic insights into local adaptation and future climate-induced vulnerability of a keystone forest tree in East Asia. Nature Communications 13.
Savolainen, O., Pyhäjärvi, T., Knürr, T., 2007. Gene Flow and Local Adaptation in Trees. Annu. Rev. Ecol. Evol. Syst. 38, 595–619. https://doi.org/10.1146/annurev.ecolsys.38.091206.095646
Schmickl, R., Yant, L., 2021. Adaptive introgression: how polyploidy reshapes gene flow landscapes. New Phytologist 230, 457–461. https://doi.org/10.1111/nph.17204
Sharik, Terry L. and Barnes, Burton V., 1971. Hybridization in Betula Alleghaniensis Britt. and B. Lenta L.: A Comparative Analysis of Controlled Crosses. Forest Science 17, 415–424.
Sharik, T.L., Barnes, B.V., 1979. Natural variation in morphology among diverse populations of yellow birch (Betula alleghaniensis) and sweet birch (B. lenta). Can. J. Bot. 57, 1932–1939. https://doi.org/10.1139/b79-242
Sharik, T.L., Barnes, B.V., 1976. Phenology of shoot growth among diverse populations of yellow birch (Betula alleghaniensis) and sweet birch (B. lenta). Can. J. Bot. 54, 2122–2129. https://doi.org/10.1139/b76-228
Sinha, D., Abid, R., Chakraborty, W., Rashid, M., Gupta, L.K., Khan, B., Datta, P.N., Noor, S., Pomila, Ghazanfar, S., Saha, U., Bhattacharya, R., Seal, S., 2024. Effect of Abiotic Stress on Terpene Biosynthesis in Plants, in: Nikalje, G.C., Shahnawaz, Mohd., Parihar, J., Qazi, H.A., Patil, V.N., Zhu, D. (Eds.), Plant Secondary Metabolites and Abiotic Stress. Wiley, pp. 481–524. https://doi.org/10.1002/9781394186457.ch16
Soltis, P.S., Soltis, D.E. (Eds.), 2012. Polyploidy and Genome Evolution. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-31442-1
Sork, V.L., Aitken, S.N., Dyer, R.J., Eckert, A.J., Legendre, P., Neale, D.B., 2013. Putting the landscape into the genomics of trees: approaches for understanding local adaptation and

population responses to changing climate. Tree Genetics & Genomes 9, 901–911. https://doi.org/10.1007/s11295-013-0596-x

Sork, V.L., Davis, F.W., Westfall, R., Flint, A., Ikegami, M., Wang, H., Grivet, D., 2010. Gene movement and genetic association with regional climate gradients in California valley oak (Quercus lobata Née) in the face of climate change. Molecular Ecology 19, 3806–3823. https://doi.org/10.1111/j.1365-294X.2010.04726.x

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., Morgenstern, B., 2006. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Research 34, W435–W439. https://doi.org/10.1093/nar/gkl200

Steane, D.A., Potts, B.M., McLean, E., Prober, S.M., Stock, W.D., Vaillancourt, R.E., Byrne, M., 2014. Genome-wide scans detect adaptation to aridity in a widespread forest tree species. Molecular Ecology 23, 2500–2513. https://doi.org/10.1111/mec.12751

Taylor, S.A., Larson, E.L., 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. Nat Ecol Evol 3, 170–177. https://doi.org/10.1038/s41559-018-0777-y

Thomson, A.M., 2013. Phylogeography, introgression, and population structure of the eastern North American birches Betula alleghaniensis, B. papyrifera, and B. lenta (Thesis (PhD)). Concordia University.

Thomson, A.M., Dick, C.W., Pascoini, A.L., Dayanandan, S., 2015. Despite introgressive hybridization, North American birches (Betula spp.) maintain strong differentiation at nuclear microsatellite loci. Tree Genetics & Genomes 11, 101. https://doi.org/10.1007/s11295-015-0922-6

Van Houtven, G., Phelan, J., Clark, C., Sabo, R.D., Buckley, J., Thomas, R.Q., Horn, K., LeDuc, S.D., 2019. Nitrogen deposition and climate change effects on tree species composition and ecosystem services for a forest cohort. Ecological Monographs 89, e01345. https://doi.org/10.1002/ecm.1345

Wang, L., Liu, J., Shen, Y., Pu, R., Hou, M., Wei, Q., Zhang, X., Li, G., Ren, H., Wu, G., 2021. Brassinosteroids synthesised by CYP85A/A1 but not CYP85A2 function via a BRI1-like receptor but not via BRI1 in Picea abies. Journal of Experimental Botany 72, 1748–1763. https://doi.org/10.1093/jxb/eraa557

Wang, T., Hamann, A., Spittlehouse, D., Carroll, C., 2016. Locally Downscaled and Spatially Customizable Climate Data for Historical and Future Periods for North America. PLoS ONE 11, e0156720. https://doi.org/10.1371/journal.pone.0156720

Wang, X., Li, J., Guo, X., Ma, Y., Qiao, Q., Guo, J., 2019. PlWRKY13: A Transcription Factor Involved in Abiotic and Biotic Stress Responses in Paeonia lactiflora. IJMS 20, 5953. https://doi.org/10.3390/ijms20235953

Xiao, Jun, Cheng, H., Li, X., Xiao, Jinghua, Xu, C., Wang, S., 2013. Rice WRKY13 regulates cross talk between abiotic and biotic stress signaling pathways by selective binding to different cis-elements. Plant Physiol 163, 1868–1882. https://doi.org/10.1104/pp.113.226019

Xing, H., Fu, X., Yang, C., Tang, X., Guo, L., Li, C., Xu, C., Luo, K., 2018. Genome-wide investigation of pentatricopeptide repeat gene family in poplar and their expression analysis in response to biotic and abiotic stresses. Sci Rep 8, 2817. https://doi.org/10.1038/s41598-018-21269-1

Ye, Y., Ding, Y., Jiang, Q., Wang, F., Sun, J., Zhu, C., 2017. The role of receptor-like protein kinases (RLKs) in abiotic stress response in plants. Plant Cell Rep 36, 235–242. https://doi.org/10.1007/s00299-016-2084-x

Zeng, Zou, Bai, Zheng, 2002. Preparation of Total DNA from “Recalcitrant Plant Taxa.” Acta Botanica Sinica 44, 694–697.

Zhang, H., Li, X., Song, R., Zhan, Z., Zhao, F., Li, Z., Jiang, D., 2022. Cap-binding complex assists RNA polymerase II transcription in plant salt stress response. Plant Cell & Environment 45, 2780–2793. https://doi.org/10.1111/pce.14388

Zhang, P., Wang, Y., Gu, X., 2020. RNA 5-Methylcytosine Controls Plant Development and Environmental Adaptation. Trends in Plant Science 25, 954–958. https://doi.org/10.1016/j.tplants.2020.07.004

Zhang, Y., Li, H., Cao, G., Dong, J., Lv, M., Su, S., Bai, Q., 2025. Heat Shock Proteins of Pistacia chinensis Could Promote Floral Development Under Drought Stress. Forests 16, 395. https://doi.org/10.3390/f16030395

Zinck, J.W.R., Rajora, O.P., 2016. Post-glacial phylogeography and evolution of a wide-ranging highly-exploited keystone forest tree, eastern white pine (Pinus strobus) in North America: single refugium, multiple routes. BMC Evol Biol 16, 56. https://doi.org/10.1186/s12862-016-0624-1

SUPPLEMENTAL INFORMATION

SUPPLEMENTAL METHODS

Supplemental Methods S1. DNA Extraction Protocol

Genomic DNA was extracted from 5-10 individuals per population using either 20 mg of silica-dried leaf tissue or entire seedlings. A modified 3% CTAB protocol based on Zeng *et al.* (2002) was used for both tissue types. Samples were mechanically homogenized with 5 mm steel balls in a TissueLyser. For seedlings, homogenization was conducted in 1 mL of CTAB-free buffer composed of 100 mM Tris-HCl (pH 8.0), 1.4 M NaCl, and 20 mM EDTA. Following homogenization, samples were incubated on ice for 10 minutes and centrifuged at 9,000 rpm. The supernatant was discarded. DNA was extracted using 800 µL of 3% CTAB buffer containing 100 mM Tris-HCl (pH 8.0), 1.4 M NaCl, 20 mM EDTA, 4% polyvinylpyrrolidone (PVP), and 0.5% β-mercaptoethanol. To this, 5 µL RNase (Nucleospin RNase, Thermo Fisher Scientific) was added, and samples were incubated at 65°C for 60 minutes in a shaking water bath.

An equal volume of chloroform:isoamyl alcohol (24:1) was added, mixed by inversion, and centrifuged at 12,000 rpm for 15 minutes at room temperature. The upper aqueous phase was transferred to a new 2 mL microcentrifuge tube, and DNA was precipitated by adding 1/2 volume of 5 M NaCl and 2/3 volume of ice-cold isopropanol. Samples were incubated at room temperature for 1 hour and then centrifuged at 10,000 rpm for 10 minutes, the supernant poured off the pelleted DNA.

DNA pellets were washed twice with 200 µL of 70% ethanol, centrifuged at 12,000 rpm for 2 minutes, and allowed to air-dry for 20-60 minutes. DNA was resuspended in 30 µL TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0).

Supplemental Methods S2: Genotyping and Bioinformatics Processing

Raw reads were quality-filtered and adapter-trimmed using cutadapt v4.6 (Martin, 2011) with parameters -e 0.2 -m 50. Reads were demultiplexed using process_radtags in STACKS v2.66 with settings to rescue barcodes and truncate reads for uniformity [-c -r -t 100 -q -s 0 --barcode_dist_1 -E phred33 --renz_1 nsiI --renz_2 mspI].

Sequencing reads were aligned to the *Betula pendula* reference genome (GCA_900184695.1) using BWA v0.7.17 (Li & Durbin, 2009), and alignments were filtered using SAMtools v1.13 (Danecek *et al.*, 2021) with the flags [-sb -q -F 4 -F 256 -F 2048]. SNPs were called using gstacks with the –max-clipped flag, and variants were exported using the populations module [-p 2 -r 0.6 --ordered-export --fasta-loci –vcf].

The initial dataset contained 337,491 SNPs across 243 individuals, with a mean coverage of 46.36x and 22.43% missing data. SNPs were filtered using 05_filter_vcf_fast.py to retain genotypes with a minimum 4x coverage, remove SNPs with <20% missing per group, and retain rare alleles in ≥3 individuals. This reduced the dataset to 31,048 SNPs. Individuals with >40% missing data were excluded (n=22), and a second filtering round increased the SNP count to 44,579. Loci with FIS > 0.8 or extreme allele balance (<0.1 or >0.8) were removed using stacks_workflow scripts 08-10 (Normandeau, n.d.). Redundant SNPs within 100 Kbp were collapsed, yielding 32,295 SNPs across 221 individuals.

Supplemental Methods S3: Paralog Detection and Genotype Weighting in polyRAD for Hexaploid Genomic Data

To account for paralogy and allele dosage in this hexaploid species, filtering was performed in polyRAD (Clark *et al.*, 2019, 2022). Read data was imported with settings to avoid SNP phasing and accommodate multiple ploidy levels [possibleploidies = list(2,6, c(2,2,2)), min.ind.with.reads = 170, min.ind.with.minor.allele = 3]. Overdispersed loci were identified using TestOverdispersion() and filtered based on a heterozygosity threshold (ploidy-1/ploidy) of >0.833. Loci were retained using SubsetByLocus(), retaining 29,674 loci. Three outlier individuals, believed to represent putative hybrids, were identified based on population structure analysis and removed. Final principal component analysis was performed using IteratePopStruct() in polyRAD with 8 PCs and a tolerance of 0.005. The first two PC axes were extracted, and the mean value per population was calculated for use in subsequent models. Weighted mean genotypes were extracted using GetWeightedMeanGenotypes(), which returns the average dosage of the VCF-defined non-reference allele per population.

Supplemental Methods S4: Partitioning of Genetic Variation

To evaluate the relative influence of environmental, geographic and neutral processes on genetic variation in *Betula alleghaniensis*, a series of redundancy analyses (RDA) and partial redundancy analyses (pRDA) were conducted using the rda() function in the *vegan* R package, following the approach of Capblancq *et al.* (2023). The response variable consisted of multivariate local allele frequencies calculated for each population using the individual allele frequencies from the GetWeightedMeanGenotypes() function in polyRAD. These frequencies,

averaged across individuals in a population, were used as the dependent matrix in all ordination models.

Climatic predictors were obtained from ClimateNA v7.21 (Wang *et al.*, 2016)) for the 1961-1990 normal period, extracted across the Betula alleghaniensis range and at the coordinates of each sampling locality. To reduce collinearity, Pearson's Pairwise correlations were calculated in base R using the cor() function. From this matrix, one variable was removed from each pair with a correlation coefficient $|r| \geq 0.7$ from the range-wide climate variables. This procedure resulted in the retention of nine uncorrelated climate variables that collectively captured the major climatic gradients across the range: annual heat-moisture index (AHM), degree-days below 0°C (DD_0), degree-days above 5°C (DD5), extreme minimum temperature (EMT), mean annual solar radiation (MAR), precipitation as snow (PAS), summer precipitation (PPT_sm), winter precipitation (PPT_wt), and mean annual relative humidity (RH).

Geographic structure was modelled using distance-based Moran's eigenvector maps (dbMEMs). A Euclidean distance matrix was first generated from the UTM coordinates of each population's centroid. The resulting distance matrix was passed to the dbmem() function in the adespatial package to compute Moran's eigenvectors. The first three positive eigenvectors (those with positive spatial autocorrelation) were retained as predictors representing spatial structure across the landscape.

Neutral genetic structure was incorporated using principal components derived from the polyRAD population structure analysis. The IteratePopStruct() function in polyRAD was used to perform a PCA on weighted mean genotypes, and the first two principal component axes were extracted. Mean PC1 and PC2 scores were calculated for each population and used as covariates

in the pRDA models to control the underlying structure unrelated to environmental or geographic factors.

A full RDA model was constructed with all three sets of explanatory variables: the nine climate predictors, the first three dbMEM axes, and the first two population structure PC axes. To partition the genetic variance and assess the unique contributions of each variable set, three partial RDA models were also performed. In the partial models, one explanatory set was used as the predictor while the other two were treated as covariates using the Condition() argument in the rda() function. The three pRDA models were: climate conditioned on geography and neutral structure; geography conditioned on climate and neutral structure; and neutral structure conditioned on climate and geography. The statistical significance of the full and partial models was evaluated using 999 permutations with the anova.cca() function. Variance partitioning was performed using the varpart() function in vegan R, allowing both unique and shared fractions of explained genetic variance to be quantified and visualized.

Supplemental Methods S5: Identification of Putatively Adaptive Loci

RDA was conducted using the rda() function in the vegan R package (Oksanen *et al.*, 2025), with the nine uncorrelated climate variables as explanatory variables and population-level allele frequencies as response variables. RDA-raw included no covariates, while RDA-x was conditioned on the first two principal components for the polyRAD population structure PCA to correct for neutral genetic structure.

GF analyses were implemented using the gradientForest R package (Ellis *et al.*, 2012). PCNM-based spatial covariates were derived from a Euclidean distance matrix of population

coordinates, and only the first half of positive PCNM axes were retained. GF models used the same nine climate predictors and were run with the following parameters [ntree = 2000, max tree depth = log2(0.368 x n / 2), corr.threshold = 0.5, nbin = 201].

To correct allele frequencies for neutral structure in GF-x, a Bayesian standardization method was implemented following Günther and Coop (2013). Allele frequencies were modelled as draws from a multivariate normal distribution with a mean vector of zero and a covariance matrix Σ. The precision matrix Ω was sampled from a Wishart distribution using a degree-of-freedom parameter equal to the number of loci present plus two. The posterior samples of Ω were inverted to obtain Σ. The model was implemented in JAGS (Plummer, 2003, 2024) using the rjags package in R. A single MCMC chains was run for 100,000 iterations, with a 5,000 iteration burn-in. A total of 190 posterior samples of the covariance matrix were retained using the coda.samples() function. Each sample matric was standardized using Cholesky decomposition to transform allele frequencies prior to fitting the GF-x model.

Since RDA and GF do not share a common test statistic, as per Capblancq *et al.* (2023) SNPs were ranked by Mahalanobis distance (RDA-raw and RDA-x) and by R^2 importance score (GF-raw and GF-x). From each test, the top 100 ranked SNPs were retained. To identify robust signals of local adaptation, only SNPs appearing in the top 100 of at least two of the four GEA approaches were retained. Overlap between methods was visualized using the VennDiagram R Package, and the intersecting set was designated as the final set of putatively adaptive loci.

Supplemental Methods S6: Modelling the Spatial Distribution of Adaptive Genomic Variation

To model the spatial distribution of climate-associated genomic variation in *Betula alleghaniensis*, an RDA was performed using the subset of putatively adaptive loci and the same nine uncorrelated climate variables previously retained for variance partitioning. The analysis was conducted using the rda() function from the vegan R package (Oksanen *et al.*, 2025). Climate variables were first standardized (centred and scaled to unit variance) across all localities to ensure equal weighting during ordination (Capblancq *et al.*, 2020). The first two axes of the RDA (RDA1 and RDA2) were retained, and the scores (loadings of each climate variable on these axes were extracted.

Composite adaptive indices were generated for each axis following the approach of Capblancq *et al.* (2020) and Steane *et al.* (2014). Specifically, the RDA1 and RDA2 loadings were multiplied by the corresponding standardized climate variables at each location and summed to produce a value representing the predicted allele frequency shift associated with each RDA axis. These indices were projected across the range of *B. alleghaniensis* using present-day standardized climate normals (1961-1990) from ClimateNA and mapped at a 1-km^2 resolution to visualize spatial patterns in adaptive genomic variation. To identify clusters of loci associated primarily with one of the two RDA axes, k-means clustering was applied using the kmeans() function in base R. Absolute values of locus scores of RDA1 and RDA2 were used as input to reflect the strength, rather than direction, of axis-specific associations. Two clusters were defined, corresponding to loci strongly associated with one axis and weakly associated with the other.

Future climate projections were obtained from ClimateNA under the 8GCM SSP5-8.5 emissions scenario (a heavy fossil-fuel development scenario) for the 2041-2070 time period

(Mahony *et al.*, 2022). The same composite method was applied using the projected climate variables and the original RDA loadings to estimate future adaptive indices across the species' range. This enabled comparison of present and future adaptive landscapes, providing a spatial framework for evaluating genomic responses to climate change.

Supplemental Methods S7: Genetic offset and adaptive capacity

To quantify the genomic change required of *Betula alleghaniensis* populations to remain aligned with future climate conditions, the genetic offset was calculated following the method of Capblancq *et al.* (2020). This measure was derived from the previously constructed RDA-based adaptive indices (RDA1 and RDA2) projected across the species' range. Genetic offset was computed as the absolute difference between the current and future adaptive index values for both RDA1 and RDA2. These values were then summed to generate a single genetic offset value for each 1-km^2 cell across the distribution of *B. alleghaniensis*. This value represents the magnitude of multivariate genomic change needed for a population to maintain local climate adaptation under the future SSP5-8.5 climate scenario.

To assess each population's capacity to respond to such change, standing genetic variation (SGV) and the population adaptive index (PAI) were calculated at each sample site. SGV was defined as the mean allele frequency variance across the putatively adaptive loci, calculated as p x q, where p is the population allele frequency and $q = 1 - p$, following the approach of Chhatre *et al.* (2019). Higher SGV reflects greater within-population genetic diversity at adaptive loci, indicating a broader capacity to respond to environmental shifts. PAI was calculated following Bonin *et al.* (2007) as the mean absolute difference between each population’s allele frequencies

and the mean allele frequencies across all sample populations for each adaptive locus. This measure captures the degree to which a population's adaptive profile diverges from the species-wide average and reflects its relative extremeness along climate-associated axes of genomic variation. Together, SGV and PAI provide complementary perspectives on adaptive capacity; SGV measures the raw potential for adaptation, while PAI highlights populations that already occupy the genetic extremes of the species' adaptive landscape (Capblancq *et al.*, 2020).

Supplemental Methods S8: Gene Function Annotation

To investigate the potential functions of climate-associated loci in *Betula alleghaniensis*, putatively adaptive SNPs were mapped to predicted genes in the *B. pendula* reference genome (GenBank GCA_900184695.1). Because the version of the genome used for read alignment was unannotated, a draft annotation was generated using AUGUSTUS v3.5.0 (Stanke *et al.*, 2006). The *Arabidopsis thaliana* gene model was used for ab initio gene prediction due to its widespread application in plant genome annotation and its compatibility with downstream functional tools. SNPs were intersected with predicted gene regions using BEDTools v2.30.0 (Quinlan & Hall, 2010). Variants located within predicted gene boundaries were annotated as genic. For SNPs located outside annotated regions, the nearest gene was assigned using the BEDTools closest function, which identifies the closest gene model based on linear genomic distance. The resulting gene models were functionally annotated using eggNOG-mapper v2 (Cantalapiedra *et al.*, 2021), which assigned gene names, biological processes, and KEGG orthology terms based on orthology predictions from the eggNOG database. This pipeline provided functional classifications for climate-associated loci based on evolutionarily conserved gene families. The gene annotation results were used to infer the potential biological roles of loci associated with climatic gradients

and assess their relevance to stress response, metabolic pathways, and transcriptional regulation in *Betula alleghaniensis*.

SUPPLEMENTAL METHODS LITERATURE CITED

AdaptWest Project., 2022. Gridded current and projected climate data for North America at 1km resolution, generated using the ClimateNA v7.30 software (T. Wang et al., 2022).

Bonin, A., Nicole, F., Pompanon, F., Miaud, C., Taberlet, P., 2007. Population Adaptive Index: a New Method to Help Measure Intraspecific Genetic Diversity and Prioritize Populations for Conservation. Conservation Biology 21, 697–708. https://doi.org/10.1111/j.1523-1739.2007.00685.x

Capblancq, T., Lachmuth, S., Fitzpatrick, M.C., Keller, S.R., 2023. From common gardens to candidate genes: exploring local adaptation to climate in red spruce. New Phytologist 237, 1590–1605. https://doi.org/10.1111/nph.18465

Capblancq, T., Morin, X., Gueguen, M., Renaud, J., Lobreaux, S., Bazin, E., 2020. Climate-associated genetic variation in Fagus sylvatica and potential responses to climate change in the French Alps. J of Evolutionary Biology 33, 783–796. https://doi.org/10.1111/jeb.13610

Chhatre, V.E., Fetter, K.C., Gougherty, A.V., Fitzpatrick, M.C., Soolanayakanahally, R.Y., Zalesny, R.S., Keller, S.R., 2019. Climatic niche predicts the landscape structure of locally adaptive standing genetic variation. https://doi.org/10.1101/817411

Clark, L.V., Lipka, A.E., Sacks, E.J., 2019. polyRAD: Genotype Calling with Uncertainty from Sequencing Data in Polyploids and Diploids. G3 Genes|Genomes|Genetics 9, 663–673. https://doi.org/10.1534/g3.118.200913

Clark, L.V., Mays, W., Lipka, A.E., Sacks, E.J., 2022. A population-level statistic for assessing Mendelian behavior of genotyping-by-sequencing data from highly duplicated genomes. BMC Bioinformatics 23, 101. https://doi.org/10.1186/s12859-022-04635-9

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. GigaScience 10, giab008. https://doi.org/10.1093/gigascience/giab008

Ellis, N., Smith, S.J., Pitcher, C.R., 2012. Gradient forests: calculating importance gradients on physical predictors. Ecology 93, 156–168. https://doi.org/10.1890/11-0252.1

Eric Normandeau, n.d. STACKS Workflow2: RADseq workflow using STACKS2 GitHub.

Günther, T., Coop, G., 2013. Robust Identification of Local Adaptation from Allele Frequencies. Genetics 195, 205–220. https://doi.org/10.1534/genetics.113.152462

Mahony, C.R., Wang, T., Hamann, A., Cannon, A.J., 2022. A global climate model ensemble for downscaled monthly climate normals over North America. Intl Journal of Climatology 42, 5871–5891. https://doi.org/10.1002/joc.7566

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet j. 17, 10. https://doi.org/10.14806/ej.17.1.200

Oksanen, J., Simpson, G., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Solymos, P., Stevens, M., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Borman, T., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M., Lahti, L., McGlinn, D., Ouellette, M., Ribeiro Cunha, E., Smith, T., Ter Braak, C., Weedon, J., 2025. vegan: Community Ecology Package. R package.

Plummer, M., 2024. rjags: Bayesian Graphical Models using MCMC. R package version 4-17.

Plummer, M., 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing. Vienna, Austria.

Steane, D.A., Potts, B.M., McLean, E., Prober, S.M., Stock, W.D., Vaillancourt, R.E., Byrne, M., 2014. Genome-wide scans detect adaptation to aridity in a widespread forest tree species. Molecular Ecology 23, 2500–2513. https://doi.org/10.1111/mec.12751

Wang, T., Hamann, A., Spittlehouse, D., Carroll, C., 2016. Locally Downscaled and Spatially Customizable Climate Data for Historical and Future Periods for North America. PLoS ONE 11, e0156720. https://doi.org/10.1371/journal.pone.0156720

Zeng, Zou, Bai, Zheng, 2002. Preparation of Total DNA from “Recalcitrant Plant Taxa.” Acta Botanica Sinica 44, 694–697.

SUPPLEMENTAL FIGURES

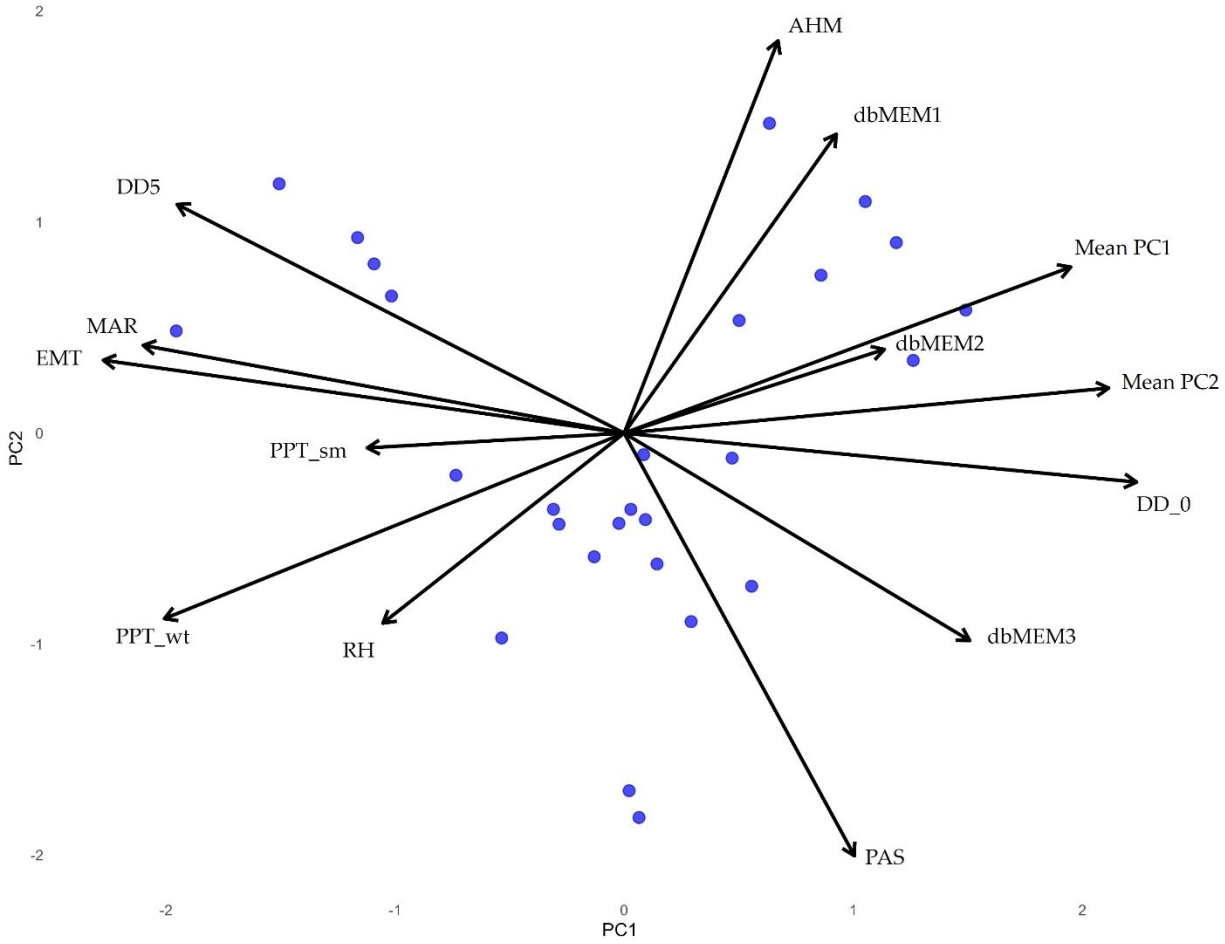


Figure S1. RDA ordination plot showing the relationship among predictor variables used in the variance partitioning analysis. Vectors represent climate variables (EMT, MAR, PPT_sm, PPT_wt, RH, PAS, DD_0, AHM, DD5), geographic structure (dbMEM axes 1-3), and neutral genetic structure (Mean PC1 and Mean PC2). Blue dots represent the 27 sampled localities.

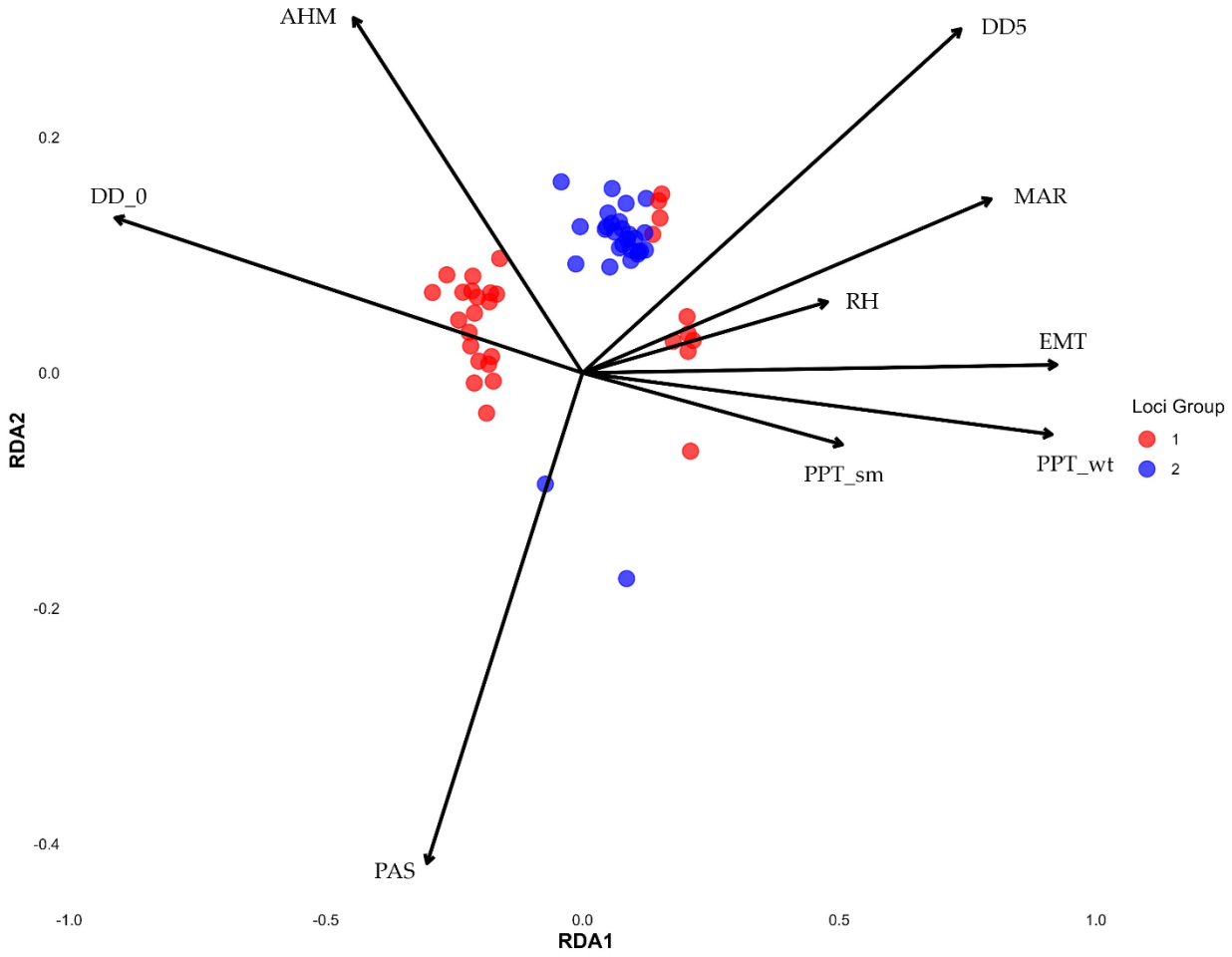


Figure S2. Adaptively enriched RDA ordination plot showing all nine climate predictors. Vectors represent the full set of climate variables used. Points represent loci identified as significantly associated with either RDA1 *or* RDA2, coloured by their respective grouping.

SUPPLEMENTAL TABLES

Table S1. The 27 sampled population locations and associated metadata for yellow birch populations included in the final analysis. Localities are identified by population codes (YB_XXX), with corresponding coordinates, site names, province or state, the number of samples retained after quality filtering and hybrid removal, and the tissue type used for DNA extraction. A total of 218 individuals were retained across 27 populations (3 to 10 individuals per population).

Locality	Latitude	Longitude	Site Name	Prov/State	Sample Count	Tissue Sampled
YB_101	44.85768	-68.629385	Penobscot	ME	10	Dried Leaf
YB_102	43.441785	-70.668625	Massabesic	ME	10	Dried Leaf
YB_103	39.05127	-79.670559	Fernow	WV	8	Dried Leaf
YB_104	41.597266	-78.773145	Kane	PA	9	Dried Leaf
YB_105	46.359099	-87.164049	Dukes	MI	10	Dried Leaf
YB_106	45.753117	-88.977224	Argonne	WI	9	Dried Leaf
YB_107	35.180891	-85.6758	Foster Falls	TN	7	Dried Leaf
YB_109	35.326663	-85.092092	North River	TN	10	Dried Leaf
YB_110	36.741569	-81.415259	Mt. Rogers	VA	10	Dried Leaf
YB_1101	45.32164	-73.09424	Gault	QC	8	Dried Leaf
YB_1102	48.273194	-89.403444	Squaretop	ON	4	Dried Leaf
YB_1103	47.341619	-84.567945	Superior	ON	9	Dried Leaf
YB_1104	46.266671	-83.415001	Thessalon	ON	5	Dried Leaf
YB_111	37.753601	-79.230005	Glenwood	VA	10	Dried Leaf
YB_112	44.053137	-71.295105	Bartlett	NH	7	Dried Leaf
YB_113	48.393144	-90.751057	Greenwood	ON	7	Dried Leaf
YB_114	47.531132	-93.470861	Marcell	MN	6	Dried Leaf

YB_115	47.115476	-93.675717	Blandin	MN	9	Dried Leaf
YB_116	43.663951	-90.574611	Kickapoo	WI	9	Dried Leaf
YB_118	45.584417	-79.21455	Algonquin	ON	5	Dried Leaf
YB_119	44.198676	-77.489004	Farm	ON	3	Dried Leaf
YB_120	42.488922	-76.774508	Finger Lakes	NY	10	Dried Leaf
YB_121	42.728327	-73.254641	Hopkins	MA	9	Dried Leaf
YB_122	35.433109	-82.733587	Pisgah	NC	10	Dried Leaf
YB_123	41.322044	-76.287615	Luzerne	PA	9	Dried Leaf
YB_2004	45.8	-64.88	Prosser Brook	NB	10	Seedling
YB_CF	44.56741	-79.70048	Copeland Forest	ON	5	Dried Leaf
			Total		218	

Table S2. Functional annotations of 49 putatively adaptive genes identified from SNPs that appeared in the top 100 ranked loci of at least two GEA methods and were significantly associated with one or both adaptive RDA axes. For each gene, the table provides a brief description, COG-based functional categories, COG-based broad category, RDA cluster assignment (based on RDA association), and the GEA methods that identified the underlying SNP(s).

Gene	Description	Function	Broad Category	Cluster	GEA Method(s)
FXXK_01000282g10936	Putative death-receptor fusion protein (DUF2428)	D - Cell cycle control, cell division, chromosome partitioning	Cellular Processes and Signaling	1	GF_raw/RDA_raw
FXXK_01000571g17561	Callose synthase	M - Cell wall/membrane/envelope biogenesis	Cellular Processes and Signaling	1	RDA_raw/RDAX
FXXK_01000128g6145	Belongs to the heat shock protein 70 family	O - Posttranslational modification, protein turnover, chaperones	Cellular Processes and Signaling	1	GF_raw/RDA_raw/ RDAX/GFX
FXXK_01000785g21037	Cysteine-rich receptor-like protein kinase	T - Signal transduction mechanisms	Cellular Processes and Signaling	1	RDA_raw/RDAX
FXXK_01001447g29159	Nodulation receptor kinase-like (SYMRK)	T - Signal transduction mechanisms	Cellular Processes and Signaling	1	RDA_raw/RDAX
FXXK_01001870g32156	G-type lectin S-receptor-like serine threonine-protein kinase	T - Signal transduction mechanisms	Cellular Processes and Signaling	1	RDA_raw/RDAX
FXXK_01000019g1086	Belongs to the adaptor complexes medium subunit family	U - Intracellular trafficking, secretion, and vesicular transport	Cellular Processes and Signaling	1	GF_raw/RDA_raw/ RDAX
FXXK_01000002g351	Pentatricopeptide repeat-containing protein	J - Translation, ribosomal structure and biogenesis	Information Storage and Processing	1	GF_raw/GFX
FXXK_01000553g17043	Belongs to the class I-like SAM-binding methyltransferase superfamily. RNA M5U methyltransferase family	J - Translation, ribosomal structure and biogenesis	Information Storage and Processing	1	RDA_raw/RDAX

Gene	Description	Function	Broad Category	Cluster	GEA Method(s)
FXXK_01000130g6235	transcription factor	K - Transcription	Information Storage and Processing	1	GF_raw/RDA_raw/ RDAX
FXXK_01001025g24403	WRKY Transcription Factor (WRKY13)	K - Transcription	Information Storage and Processing	1	GF_raw/GFX
FXXK_01000390g13280	Aldehyde dehydrogenase family	C - Energy production and conversion	Metabolism	1	RDA_raw/RDAX
FXXK_01000997g24103	NADPH adrenodoxin oxidoreductase, mitochondrial	C - Energy production and conversion	Metabolism	1	GF_raw/GFX
FXXK_01000023g1418	May be involved in modulation of pathogen defense and leaf cell death	G - Carbohydrate transport and metabolism	Metabolism	1	RDA_raw/RDAX
FXXK_01000550g16998	CRAL/TRIO, N-terminal domain	I - Lipid transport and metabolism	Metabolism	1	RDA_raw/RDAX
FXXK_01000026g1595	Domain of unknown function (DUF966)	P - Inorganic ion transport and metabolism	Metabolism	1	RDA_raw/RDAX
FXXK_01000230g9539	Potassium channel	P - Inorganic ion transport and metabolism	Metabolism	1	RDA_raw/RDAX
FXXK_01000579g17807	Belongs to the iron ascorbate-dependent oxidoreductase family	Q - Secondary metabolites biosynthesis, transport and catabolism	Metabolism	1	RDA_raw/RDAX
FXXK_01000714g19958	glycosyltransferase At5g03795	G - Carbohydrate transport and metabolism; M - Cell wall/membrane/envelope biogenesis; W - Extracellular structures	Metabolism; Cellular Processes and Signaling	1	RDA_raw/RDAX
FXXK_01000046g2870	jasmonate-zim-domain protein	S - Function unknown	Poorly Characterized	1	RDA_raw/RDAX
FXXK_01000302g11436	zinc finger CCCH domain-containing protein	S - Function unknown	Poorly Characterized	1	GF_raw/RDA_raw/ RDAX/GFX

Gene	Description	Function	Broad Category	Cluster	GEA Method(s)
FXXK_01000606g18350	Nuclear pore complex protein	S - Function unknown	Poorly Characterized	1	RDA_raw/RDAX
FXXK_01000888g22625	2-alkenal reductase (NADP( )-dependent)-like	S - Function unknown	Poorly Characterized	1	RDA_raw/RDAX
FXXK_01001207g26826	Dirigent protein metabolism	S - Function unknown	Poorly Characterized	1	RDA_raw/RDAX
FXXK_01000558g17144	Serine threonine-protein kinase	D - Cell cycle control, cell division, chromosome partitioning; T - Signal transduction mechanisms	Cellular Processes and Signaling	2	RDA_raw/RDAX
FXXK_01000384g13219	Ring finger	O - Posttranslational modification, protein turnover, chaperones	Cellular Processes and Signaling	2	RDA_raw/RDAX
FXXK_01000037g2199	isoform X1	T - Signal transduction mechanisms	Cellular Processes and Signaling	2	RDA_raw/RDAX
FXXK_01001540g29979	serine threonine-protein phosphatase	T - Signal transduction mechanisms	Cellular Processes and Signaling	2	RDA_raw/RDAX
FXXK_01002079g33341	TMV resistance protein N-like	T - Signal transduction mechanisms	Cellular Processes and Signaling	2	RDA_raw/RDAX
FXXK_01002545g35112	mitogen-activated protein kinase	T - Signal transduction mechanisms	Cellular Processes and Signaling	2	RDA_raw/RDAX
FXXK_01000167g7985	60s ribosomal protein	J - Translation, ribosomal structure and biogenesis	Information Storage and Processing	2	RDA_raw/RDAX
FXXK_01000191g8655	RNA polymerase II C-terminal domain phosphatase-like	K - Transcription	Information Storage and Processing	2	RDA_raw/RDAX
FXXK_01000023g1418	May be involved in modulation of pathogen defense and leaf cell death	G - Carbohydrate transport and metabolism	Metabolism	2	RDA_raw/RDAX

Gene	Description	Function	Broad Category	Cluster	GEA Method(s)
FXXK_01000281g10869	Glycosyl hydrolase family 10 protein	G - Carbohydrate transport and metabolism	Metabolism	2	RDA_raw/RDAX
FXXK_01001893g32326	Belongs to the FPP GGPP synthase family (GPPS)	H - Coenzyme transport and metabolism	Metabolism	2	RDA_raw/RDAX
FXXK_01000116g5544	ADP,ATP carrier protein	P - Inorganic ion transport and metabolism	Metabolism	2	RDA_raw/RDAX
FXXK_01001087g25280	Dehydrogenase reductase SDR family member	Q - Secondary metabolites biosynthesis, transport and catabolism	Metabolism	2	RDA_raw/RDAX
FXXK_01000037g2230	acyl-CoA--sterol O-acyltransferase 1-like	S - Function unknown	Poorly Characterized	2	RDA_raw/RDAX
FXXK_01000382g13178	atrnl,rnl	S - Function unknown	Poorly Characterized	2	RDA_raw/RDAX
FXXK_01000578g17748	expansin-A11-like	S - Function unknown	Poorly Characterized	2	RDA_raw/RDAX
FXXK_01001071g25082	Plant mobile domain	S - Function unknown	Poorly Characterized	2	RDA_raw/RDAX
FXXK_01001472g29386	Sieve element occlusion (SEOa)	S - Function unknown	Poorly Characterized	2	RDA_raw/RDAX
FXXK_01000142g6798	G-type lectin S-receptor-like serine threonine-protein kinase	T - Signal transduction mechanisms	Cellular Processes and Signaling	1&2	RDA_raw/RDAX
FXXK_01001025g24401	Belongs to the protein kinase superfamily. Ser Thr protein kinase family. (BRI1)	T - Signal transduction mechanisms	Cellular Processes and Signaling	1&2	GF_raw/RDA_raw/ RDAX/GFX
FXXK_01001855g32088	Cysteine-rich receptor-like protein kinase	T - Signal transduction mechanisms	Cellular Processes and Signaling	1&2	RDA_raw/RDAX
FXXK_01000733g20262	BEACH domain-containing protein	U - Intracellular trafficking, secretion, and vesicular transport	Cellular Processes and Signaling	1&2	GF_raw/RDA_raw/ RDAX/GFX
FXXK_01000163g7829	Ethylene-responsive transcription factor	K - Transcription	Information Storage and Processing	1&2	GF_raw/GFX

Gene	Description	Function	Broad Category	Cluster	GEA Method(s)
FXXK_01000550g17009	Protein of unknown function (DUF674)	L - Replication, recombination and repair	Information Storage and Processing	1&2	RDA_raw/RDAX
FXXK_01001925g32512	Cation H( ) antiporter	P - Inorganic ion transport and metabolism	Metabolism	1&2	RDA_raw/RDAX