The genomic basis of local adaptation in the white x Engelmann spruce hybrid zone

Gabriele Nocchi[1], Janek Sendrowski[2], Andy Shi[1,3], Brianne Boufford[4], Manuel Lamothe[5], Nathalie Isabel[5], Samuel Yeaman[1]

1. Department of Biological Sciences, University of Calgary, Calgary, AB, Canada
2. Bioinformatics Research Center (BiRC), Aarhus University, Aarhus, Denmark
3. Western Canada High School, Calgary, AB, Canada
4. Department of Forest Resources Management, University of British Columbia, Vancouver, BC, Canada
5. Natural Resources Canada, Canadian Forest Service, Laurentian Forestry Centre, Québec City, QC, Canada

Corresponding author: Samuel Yeaman (samuel.yeaman@ucalgary.ca)

Author contributions: GN analysed the data and wrote the manuscript. JS performed sampling and collected cambium for DNA extraction. AS helped assess snow cover data. BB generated snow cover data. ML and NI contributed to the development of genomic resources. SY conceived the project and secured funding, designed sampling and sequencing methods, provided overall supervision, and helped with writing the manuscript.

Abstract (250 words)

Hybridization between species can occur along repeated zones of contact, providing a powerful natural laboratory for studying the interplay between migration and selection, and for identifying loci involved in adaptation and reproductive isolation. However, interpreting how evolutionary processes shape genomic patterns can be challenging: repeatability of genotype-environment association alone is not strong evidence for selection, as such patterns could also occur at loci that originally differentiated between the parental species

due to drift. On the other hand, repeated directional patterns of introgression, where an allele typical of one species moves far into the range of the other, are unlikely to occur due to drift and therefore provide stronger evidence of natural selection. Here we compared hybridization and local adaptation patterns between two replicated regions within the western Canada interior spruce hybrid zone: broad latitudinal transects with gradual environmental variation and narrow elevational transects with steep environmental contrasts. Despite differences in geography and selective pressures, the two regions revealed strikingly similar genome-wide patterns of differentiation and adaptation. Genomic clines analyses confirmed this trend, uncovering consistent patterns of directional introgression between transect types. The strong similarity in differentiation, environmental adaptation and directional introgression between replicated hybrid zones is unlikely under an entirely drift-driven process, suggesting that selection plays an important role in shaping hybridization patterns between *Picea glauca* and *P. engelmannii*. Consistent with hybridization theory, we found longer introgressed genomic tracts in the elevational transects, likely because steep environmental gradients over short distances restrict recombination preserving larger ancestry blocks in the genome.

Introduction

Introgression, which is the incorporation of alleles from one species or population into the gene pool of another through hybridization and repeated backcrossing, is a common and long-recognized phenomenon in plants (Anderson 1953; Haselhorst and Buerkle 2013; Harrison and Larson 2014), and it has also been documented in many animal species in recent years (Hedrick 2013; Aguillon et al. 2022). Hybridization and subsequent introgression can have substantial impacts on adaptation, evolvability, and speciation that can vary in the direction of their effect depending on the interaction between genetics and ecology. On one hand, hybridization may introduce genetic variants that enhance the fitness

of the recipient population (adaptive introgression), may increase standing genetic diversity boosting a population's adaptive potential to environmental changes, and may also lead to speciation if hybrids become reproductively isolated from parental populations (Rieseberg et al. 2003; Baack and Rieseberg 2007; Becker et al. 2013; Brennan et al. 2014; Hamilton and Miller 2015; Mallet 2015). On the other hand, hybridization may counteract local adaptation with the introduction of maladapted alleles, may lead to unfit or unviable hybrids, and may homogenize differences between populations with incomplete reproductive isolation counteracting speciation (Barton and Hewitt 1985). In extreme circumstances, hybridization may also lead to extinction through the complete genetic assimilation of a rare species by a more common congener, such as in the case of *Argyranthemum coronopifolium* in Tenerife (Canary Islands), which was effectively replaced by hybrids and *A. frutescens* within just a few decades from the initial contact (Levin et al. 1996). The impact of introgression therefore depends upon how the hybridizing lineages are each adapted to their local environments, and the extent to which they each carry deleterious load (that could be complemented and relieved by introgression) and globally beneficial alleles (that could introgress by a selective sweep).

Studying patterns of the genetic basis of introgression across replicated hybrid zones can provide insights into the importance of selection vs. drift (Rieseberg et al. 1999; Simon et al. 2020; Westram et al. 2021; McFarlane et al. 2024; Semenov et al. 2025): if similar genomic regions exhibit similar patterns of introgression across multiple hybrid zones, this can suggest a consistent role for selection rather than purely stochastic processes. Outside of hybrid zones, shared selection pressures across repeated environmental gradients should lead to similar genotype–environment associations (GEAs), whereas drift will result in differing patterns of association, so repeated GEA signals are commonly used as evidence of selection (Whiting et al. 2024). However, if two previously isolated species come into repeated contact across multiple hybrid zones, loci that originally differentiated by genetic drift could still be repeatedly associated with the overall environmental gradient that

distinguishes the parental species. Thus, repeatability in GEA signals alone across multiple hybrid zones is not sufficient evidence to implicate natural selection (Lotterhos and Whitlock 2014, 2015; Forester et al. 2018; Leigh et al. 2021; Booker et al. 2023). However, if the regions of the genome that exhibit strong GEA signals also exhibit abnormal introgression in the same pattern across multiple hybrid zones, this would be unlikely to occur due to genetic drift, and would implicate natural selection (Rieseberg et al. 1999; Buerklen and Lexer 2008; Gompert and Buerkle 2009; Bierne et al. 2013; Simon et al. 2020; Westram et al. 2021; Langdon et al. 2022). Abnormal introgression can be inferred as a deviation of allele frequency change at a locus from the average genome-wide patterns, in the form of either directional introgression farther into one species background or the other, or particularly broad or narrow transitions in allele frequency across the hybrid zone (Gompert and Buerkle 2009, 2010, 2011, 2012). Our aim is to assess whether genomic differentiation, GEAs, and introgression patterns are repeatable between the elevational and latitudinal hybrid zones, to understand the extent to which hybridization outcomes in spruce are shaped by selection versus drift.

White spruce (*Picea glauca*) and Engelmann spruce (*Picea engelmannii*) are wind-pollinated, long-lived commercially and aesthetically important forest tree species in Canada (Rajora and Dancik 2000; De La Torre et al. 2013). White spruce is a subarctic boreal species that can tolerate both low winter temperatures as well as summer drought but is generally restricted to low elevations (Rajora and Dancik 2000; De La Torre et al. 2013). Engelmann spruce is a high-altitude subalpine forest tree with a more westerly and southerly distribution than white spruce, extending from Southern British Columbia and Alberta in the north, up to Arizona and New Mexico in the south (Rajora and Dancik 2000; De La Torre et al. 2013). Engelmann spruce exhibits low tolerance to elevated temperatures and drought conditions, inhabits wetter environments and can withstand short growing season and high snowfall (Rajora and Dancik 2000; De La Torre et al. 2013). White and Engelmann spruce are closely related species and are thought to have diverged in allopatry during the Pliocene

(Lockwood et al. 2013; De La Torre et al. 2014). A recent study proposed that their most recent contacts are likely to have started in the Rocky Mountains of southern Colorado and Wyoming at the last glacial maximum (LGM), when both species were displaced south of their current distribution. After the glacial retreat, both species started re-expanding, causing the hybrid zone to gradually shift northwest, reaching British Columbia approximately 14,000 years ago and continuing to expand until reaching its current extent in the late Holocene (De La Torre et al. 2014).

Today, these two species hybridize extensively where their ranges overlap in western Alberta and British Columbia, so that the complex of these two species and their hybrids is commonly referred to as “interior spruce” (Rajora and Dancik 2000). As a result, it can be difficult to distinguish these two species by observation except by the morphology of their cones (Taylor 1959; Rajora and Dancik 2000). Hybrids display clinal genetic variation across the hybrid zone and tend to occupy habitats that are ecologically and elevationally intermediate between those of the parental species, suggesting that the interior spruce hybrid zone in western Canada is maintained according to the bounded hybrid superiority model (Horton et al. 1959; Rajora and Dancik 2000; Ledig et al. 2006; De La Torre et al. 2013, 2014). Local adaptation is well supported by common garden experiments as well as previous genomic studies, which have also revealed strong differentiation between *P. glauca* and *P. engelmannii* (De La Torre et al. 2013, 2014; Hornoy et al. 2015; Yeaman et al. 2016; Capblancq et al. 2022). However, the axis of introgression aligns with climatic gradients, making it unclear how much of the genomic divergence reflects a history of selection versus neutral drift in the parental species prior to secondary contact (Horton et al. 1959; Rajora and Dancik 2000; Ledig et al. 2006; De La Torre et al. 2013, 2014).

Here, to disentangle the importance of drift and selection in shaping the interior spruce hybrid zone, we compare genomic patterns across two datasets: a previously studied broad latitudinal gradient spanning >1000km throughout Alberta and British Columbia (Fig. 1a) (Yeaman et al. 2016), and a new dataset collected across a narrower elevational zone

133 spanning ~100km in the Canadian Rocky Mountains (Fig. 1c). To characterize hybridization
134 across this elevational gradient, we sampled interior spruce individuals along multiple
135 transects spanning a system of valleys within the Rocky Mountains, covering this narrow
136 region of the hybrid zone (Fig. 1c). At seven locations along these valleys, we collected
137 samples from three distinct environments: valley bottoms (<1600 m), and high elevation
138 north- and south-facing slopes (>2000 m). Based on casual observations, forests on high
139 north-facing slopes tend to have moister soil and hold snow longer in the springtime than
140 those on high south-facing slopes and valley bottoms, so we expected this might generate
141 some fine-grained natural selection within each transect.

142 To characterize how the environment varies between these species, we identify a set of
143 environmental variables that clearly differentiate their ecological niches and capture this fine-
144 scale within-transect variation. We then use these variables in GEA analyses to identify the
145 genomic regions most strongly associated with the main axis of environmental niche
146 divergence. This analysis should detect candidates driving local adaptation in the parental
147 species, but it also may pick up any regions of the genome that happened to evolve strong
148 differentiation due to drift or intrinsic reproductive isolating mechanisms. To further study the
149 potential importance of natural selection, we use a genomic clines analysis to characterize
150 the patterning of introgression at each locus, relative to average patterns across the genome
151 (Gompert and Buerkle 2009, 2010, 2011, 2012). If consistent patterns of directional
152 introgression are observed across the latitudinal and elevational transects, this would further
153 support the hypothesis that selection plays a key role in shaping hybridization between white
154 and Engelmann spruce. In this case, the genomic regions showing repeated directional
155 introgression would constitute strong candidates for adaptive introgression (Rieseberg et al.
156 1999; Buerklen and Lexer 2008; Gompert and Buerkle 2009; Bierne et al. 2013; Simon et al.
157 2020; Westram et al. 2021; Langdon et al. 2022).

158 To further explore how the geography of hybridization impacts signatures in the genome, we
159 conduct a comparative analysis of the chromosomal patterning of loci with strong signals of

160 association to environment. Given that migration rates are likely similar across the species'
161 range, the elevational and latitudinal datasets represent distinct relationships between
162 migration rate and spatial distance across the hybrid zone. The elevational transects
163 encompass strong environmental variation over shorter spatial scales, resulting in stronger
164 variation in selection over a shorter distance (and therefore, higher migration). As a result,
165 there should be fewer opportunities for recombination to break up the patterns of linkage
166 disequilibrium (LD) from the parentals. If this difference in the geography of hybridization
167 matters, we should observe larger genomic blocks surrounding selected loci in the
168 elevational transects compared to the latitudinal transects (Barton 1983; Sedghifar et al.
169 2016; Shchur et al. 2020). We test these predictions by computing the spatial-autocorrelation
170 along the chromosomes in both F_{ST} and p-values from the GEA analysis, and comparing
171 values between the elevational and latitudinal transects.

172

173 # Results

174 **Super-scaffold assembly**

175 While the reference genome assembly of *Picea glauca* is highly fragmented (Warren et al.
176 2015), we used a previously constructed linkage map based on cDNA sequences to position
177 contigs onto chromosome scale scaffolds, to facilitate study of genome-wide patterns. To
178 maximize confidence, we identified the best scaffold match for each cDNA based on
179 sequence coverage and identity and discarded unmapped cDNA as well as those with low-
180 quality mapping or unresolved multi-mapping (see *Methods* for details). This process
181 retained 21,252 cDNAs, of which 11,914 were included in the available *P. glauca* linkage
182 map, including 14,727 genes spread across 12 linkage groups (LGs) (Pavy et al. 2017;
183 Gagalova et al. 2022). These 11,914 cDNAs mapped to 10,475 unique scaffolds of the *P.*
184 *glauca* reference genome, which we ordered along the 12 linkage groups according to the
185 cDNAs centimorgan (cM) positions in the linkage map. The resulting super-scaffold

assembly spanned 3,426,128,031 bp (~3.42 Gbp), representing approximately 13% of the WS77111v2 *P. glauca* reference genome assembly (~26.5 Gbp). The final size of our assembly is comparable to a recently published super-scaffold assembly for *P. glauca*, which achieved a size of ~3.86 Gbp by using a less conservative cDNA mapping approach (Gagalova et al. 2022).

Sampling and SNP Discovery

We sampled 331 spruce individuals, including white spruce (*Picea glauca*), Engelmann spruce (*Picea engelmannii*), and their hybrids, from 21 populations within a narrow section of the western Canada interior spruce hybrid zone which extends from central British Columbia (B.C.) into western Alberta (Fig. 1c). Samples were collected from seven transects along the Trans-Canada Highway going through Canmore, Banff, and Yoho. Each transect included three populations, with 14 to 16 individuals sampled per population (supplementary fig. S1, Supplementary Material online). For each transect, we sampled a north-facing high-altitude population (N), a south-facing high-altitude population (S), and a valley population (V) (Fig. 1c; supplementary fig. S2, Supplementary Material online). Using a target sequencing approach with a 10 Mb target, we identified a set of 119,664 SNPs (MAF > 0.05) following post-processing (see *Methods* for details). This set included only SNPs located on the scaffolds of the *P. glauca* reference genome that we were able to anchor to the *P. glauca* linkage map (Gagalova et al. 2022) in our super-scaffold assembly construction. We termed this dataset the "elevational dataset".

Additionally, we gathered available exome capture data for 575 spruce individuals from 249 populations, derived from previous studies (Suren et al. 2016; Yeaman et al. 2016). This dataset spans both species' geographic ranges in B.C. and Alberta and encompasses the entire western Canada interior spruce hybrid zone, with the exclusion of areas within national parks (Fig. 1a; supplementary fig. S1, Supplementary Material online). After

212 applying the same SNP-calling pipeline and post-processing as for the elevational dataset to
213 ensure consistency, we identified a set of 265,655 SNPs (MAF > 0.05) located on the
214 scaffolds of our super-scaffold assembly. We termed this dataset the "latitudinal dataset".

215

216 **Population structure**

217 We analysed genome-wide variation in both elevational and latitudinal transects using
218 principal component analysis (PCA) and a variational Bayesian framework implemented in
219 *fastSTRUCTURE* (Raj et al. 2014). In both datasets, *fastSTRUCTURE* identified K=2 as the
220 number of clusters that maximizes the log marginal likelihood (K*ε) of the data (Fig. 1b and
221 d; supplementary fig. S3, Supplementary Material online). Concordantly, PCA's first principal
222 component (PC1) revealed two main clusters of individuals present in both datasets, while
223 capturing a large portion of the variation in the data (70% in the latitudinal dataset and 63%
224 in the elevational dataset) as indicated by the sharp drop in eigenvalue scores after PC1
225 (supplementary figs. S4 and S5, Supplementary Material online).

226 To assign the two clusters identified with *fastSTRUCTURE* and PCA to their respective
227 species, we added publicly available genomic data of a pure individual of *P. glauca* and one
228 of *P. engelmannii* to each dataset, retrieved from the European Nucleotide Archive (ENA)
229 (see *Methods* for details). By repeating the *fastSTRUCTURE* analysis at K=2, we could
230 clearly identify which species ancestry each cluster represented based on the pure
231 individuals admixture coefficients (supplementary fig. S6, Supplementary Material online).
232 Overall, PCA and *fastSTRUCTURE* effectively distinguished the different genomic
233 background of the two spruce species and produced highly concordant results: the
234 distribution of spruce individuals along PC1 largely reflected the arrangement of the
235 individuals based on the ancestry coefficients estimated with *fastSTRUCTURE* at K=2
236 (supplementary fig. S7, Supplementary Material online). In PCA, the remaining principal
237 components (PC2, PC3) explained considerably less variation than PC1 in both the

elevational and latitudinal transects and separated a few second- and third-degree relatives found within some populations from the rest of the individuals (supplementary figs. S4 and S5, Supplementary Material online). Overall, genomic relatedness levels were low according to our estimation, with only two potential half-sibling relationships (kin ~0.25) in the elevational dataset and ten in the latitudinal dataset (supplementary fig. S8, Supplementary Material online). Both datasets showed few third-degree relationships (supplementary fig. S8, Supplementary Material online).

Notably, individuals in the elevational transects exhibited patterns of introgression strongly reflecting their environmental context (Fig. 2; supplementary fig. S9, Supplementary Material online). Within each sampled environment (N, S, V), we tend to find a transition from Engelmann-dominant ancestry in the north-western transects to white spruce-dominant ancestry in the south-eastern transects, but this shift occurs within different environment types over different regions of the hybrid zone. For the furthest north-west transects, the shift occurs in the valleys, for transects around the middle of the sampled range the transition occurs on high S-facing slopes, and for transects furthest to the south-east the transition occurs on high N-facing slopes (Fig. 2; supplementary fig. S9, Supplementary Material online). Thus, while individual transects ancestry tend to be more Engelmann dominated in the north-west and white-dominated in the south-east, the ancestry distribution shifts across environment types, with these shifts becoming particularly pronounced in the transects located in the middle of the sampling range.

For the latitudinal dataset, ancestry distribution aligned well with the geographical ranges of the two species in western Canada (Fig. 1a) (Rajora and Dancik 2000; De La Torre et al. 2013, 2014; Gagalova et al. 2022). Potentially pure *P. glauca* and *P. glauca*-dominant hybrids were found predominant in the northern and eastern part of the sampling range going into Alberta, while *P. engelmannii* ancestry appeared more prevalent in south-western British Columbia. More balanced species density with varying admixture levels were observed in central and western B.C. and along the southern B.C.-Alberta border, matching

265 with the extent of what is commonly referred to as the western Canada interior spruce hybrid
266 zone (Fig. 1a).

267

268 **Population differentiation**

269 We measured the extent of genetic differentiation between populations in each dataset using
270 the F_{ST} statistic on a per-SNP basis, implemented in *VCFtools* (Danecek et al. 2011). In the
271 elevational dataset, mean F_{ST} between the 21 populations sampled was 0.03 (median =
272 0.01, sd = 0.05), based on 119,664 SNPs (supplementary fig. S10, Supplementary Material
273 online). This varied from 0.025 to 0.042 when assessed by chromosome. In the latitudinal
274 dataset, mean F_{ST} between 249 populations was 0.07 (median=0.6, sd=0.07), based on
275 265,655 SNPs (supplementary fig. S10, Supplementary Material online). This varied from
276 0.067 to 0.084 when assessed by chromosome. For reference, mean F_{ST} between *P. glauca*
277 and *P. engelmannii*, estimated by selecting pure species individuals based on the
278 *fastSTRUCTURE* admixture coefficient Q ($Q \geq 0.95$), was 0.10 (median=0.04, sd=0.15) in
279 the elevational dataset and 0.14 (median=0.07, sd=0.17) in the latitudinal dataset
280 (supplementary fig. S10, Supplementary Material online).

281 Comparing the SNPs analysed in each dataset, 11,990 were shared between the two
282 datasets. This shared set was limited, in part due to small population sizes and missing data
283 in the latitudinal dataset which prevented F_{ST} computation for several loci, but primarily due
284 to differences in the sequence capture design used for each dataset. For those regions of
285 the genome represented in both panels, the correlation between the F_{ST} values for the
286 shared SNPs was high (Pearson's correlation = 0.85, $p < 2.2e-16$; Kendall's tau = 0.48, $p <$
287 $2.2e-16$), showing little variation between linkage groups (Fig. 3b and d).

288 Overall, the F_{ST} landscape between the elevational and latitudinal transects appears highly
289 repeated, with a visually concordant distribution of peaks and valleys in the resulting
290 Manhattan plots (Fig. 3a and c; supplementary fig. S11, Supplementary Material online).

291 Furthermore, to compare the F_{ST} landscape architecture between the two datasets and the
292 extent of any signatures of clustering, we calculated Moran's I index of spatial
293 autocorrelation on chromosomal distance using F_{ST} values from the 11,990 shared SNPs for
294 each scaffold and linkage group (see *Methods* for details). Although not statistically
295 significant (scaffolds paired t-test p-value = 0.49, n = 27; LGs paired t-test p-value = 0.22; n
296 = 12), this analysis showed that F_{ST} chromosomal autocorrelation tended to be slightly
297 higher in the elevational than in the latitudinal transects when assessed by LG (Elevational
298 scaffolds Moran's I – Latitudinal scaffolds Moran's I mean and median: 0.007 and 0.005;
299 Elevational LGs Moran's I – Latitudinal LGs Moran's I mean and median: 0.006 and 0.002)
300 (supplementary fig. S12, Supplementary Material online). However, the sample size of only
301 12 LGs doesn't allow for much power, and no strong trend was detected in the scaffolds
302 (supplementary fig. S12, Supplementary Material online).

303

304 **Genome-environment association analysis**

305 To identify the particular combination of environments that covaries most strongly with
306 variation in ancestry, we performed redundancy analysis (RDA) by regressing the individual
307 trees' hybrid indexes, computed with *fastSTRUCTURE*, against a set of environmental
308 variables considered important in describing the niches of the two parental species. This
309 approach involves a specific type of multiple regression aimed at identifying the linear
310 combination of environmental variables that accounts for the largest portion of variation in
311 the hybrid index, and therefore best explains the introgression patterns that we observed.
312 RDA1 (the first axis of RDA variation i.e. the linear combination of predictors explaining most
313 hybrid index variance) was then used as a predictor in genome-environment association
314 analyses (GEAs). For the elevational dataset, we included six weakly correlated (Pearson's
315 correlation coefficient < 0.7) environmental variables: aspect (the compass direction that a
316 terrain faces), slope (the steepness of the terrain), air temperature at 2 meters, diffuse
317 horizontal irradiation (DIF), photovoltaic electricity output (PVOUT), and snow melt day (see

318 *Methods* for details). For the latitudinal dataset, we used the same variables except snow
319 melt day. The latter, being high-resolution data, was not feasible to generate for the vast
320 geographic area covered by the latitudinal dataset.

321 In the elevational dataset, RDA1 explained 73% of the variation in the hybrid index, with
322 snow melt day and PVOUT being the most important variables, followed by air temperature
323 at 2 meters, slope, aspect, and DIF (supplementary fig. S13, Supplementary Material
324 online). In contrast, in the latitudinal dataset RDA1 explained 56% of the total hybrid index
325 variation. The most influential variables were slope and PVOUT, followed by air temperature
326 at 2 meters and DIF (supplementary fig. S13, Supplementary Material online). In this
327 dataset, aspect contributed negligibly to the introgression patterns observed across latitude,
328 in strong contrast with its contribution in the elevational dataset (supplementary fig. S13,
329 Supplementary Material online).

330 We used the resulting RDA1 axis as predictors in the GEA, conducting separate analyses for
331 the elevational and latitudinal datasets. The elevational dataset analysis included 119,664
332 SNPs, while the latitudinal included 265,655 SNPs (MAF > 0.05). For each SNP, we
333 performed a likelihood-ratio test comparing two models: a null model including only the
334 intercept, where the SNP frequency was predicted using only the mean frequency; and a
335 predictor model which used RDA1 as the environmental predictor to infer SNP frequency.
336 The likelihood-ratio test was used to compare the goodness of fit between the predictor
337 model and the intercept-only model. A significantly better fit by the predictor model results in
338 a low p-value, reflecting the predictor's importance. Conversely, if the null model is
339 supported by the observed data, the likelihoods of the two models would not differ beyond
340 the sampling error, resulting in higher p-values. Similar to the F_{ST} landscape, the adaptive
341 landscape between the elevational and latitudinal transects appeared highly repeated, with a
342 visually strikingly concordant distribution of peaks and valleys in the Manhattan plots derived
343 from the two GEAs (Fig. 4a and c; supplementary fig. S14, Supplementary Material online).

Between the SNPs analysed in each dataset, 31,667 were shared, exhibiting very high correlation in p-values between the two GEA analyses (Pearson's correlation: 0.87, $p < 2.2e-16$; Kendall's tau = 0.62, $p < 2.2e-16$), with little variation between linkage groups (Fig.4b; supplementary fig. S15, Supplementary Material online).

As above, to compare the architecture of the genomic landscape between the two datasets and the extent of any signatures of clustering, we calculated Moran's I index of spatial autocorrelation using GEA p-values from the 31,667 shared SNPs for each scaffold and linkage group (see *Methods* for details). Consistent with our expectation based on the theory of hybridization and migration-selection balance, the levels of GEA p-values chromosomal autocorrelation within scaffolds and LGs were significantly ($p < 0.05$) higher (scaffolds paired t-test p-value = 0.02, n = 82; LGs paired t-test p-value = 0.01, n = 12) in the elevational rather than latitudinal dataset (Elevational scaffolds Moran's I – Latitudinal scaffolds Moran's I mean and median: 0.02 and 0.0005; Elevational LGs Moran's I – Latitudinal LGs Moran's I mean and median: 0.01 and 0.01) (supplementary fig. S16, Supplementary Material online).

Furthermore, results of a secondary log-likelihood test-based GEA incorporating population structure correction using PC1 from our PCA analysis (see *Methods* for details) showed that the resulting p-values appear entirely uncorrelated between the elevational and latitudinal analyses, in strong contrast with those of our GEA results without population structure correction (supplementary fig. S17, Supplementary Material online). This suggests that GEA methods that correct for population structure may not be suitable to use when there is strong confounding between genetic structure and environmental variation, as previously reported (Lotterhos 2019; Booker et al. 2023). This issue is especially pronounced in the elevational dataset, where this type of confounding is maximized due to our targeted sampling of environments that covary strongly with the hybrid index (Fig. 2).

370 **Patterns of introgression in the western Canada interior spruce hybrid zone: genomic**
371 **clines analysis**

372 Genomic clines analysis is used to infer the movement of ancestry blocks into different
373 genomic backgrounds and to assess patterns of introgression in hybrid zones, by modelling
374 the ancestry probabilities of SNPs as a function of the hybrid index. Here, we aim to identify
375 regions of the genome exhibiting extreme patterns of introgression, which might occur if
376 some loci are more strongly introgressed into one species or the other due to their adaptive
377 potential (excess ancestry), or if hybridization at some loci results in significantly improved
378 fitness such as through complementation of deleterious recessive alleles or, conversely, lead
379 to a significant fitness reduction (i.e. due to reproductive incompatibilities) (Gompert and
380 Buerkle 2009, 2010, 2011, 2012). We applied genomic clines analyses to assess the
381 patterns of introgression between *P. glauca* and *P. engelmannii* in the western Canada
382 interior spruce hybrid zone. We fit genomic clines using two approaches: a Bayesian
383 genomic clines model implemented in *BGC* (Gompert and Buerkle 2012) and a logistic
384 regression-based model implemented in *introgress* (Gompert and Buerkle 2010). We used
385 *introgress* to have a source of comparison and verify the consistency of *BGC* but only results
386 from the latter approach were analysed further (see *Methods* for details).

387 In order to define parental pure species populations for the genomic clines model, we
388 classified the sampled individuals as either *P. glauca*, *P. engelmannii* or hybrids according to
389 the admixture coefficient computed with *fastSTRUCTURE* (see *Methods* for details). We
390 performed two separate genomic clines analyses, one for the elevational and one for the
391 latitudinal dataset respectively and included in the models only SNPs with an allele
392 frequency differential between parental “pure” species populations greater than 0.6. Limiting
393 the number of SNPs tested in *BGC* was necessary, as using the full SNP set would lead to
394 infeasible run times due to the computational demands of its Bayesian framework.
395 Therefore, we opted to include only ancestry-informative SNPs (those with substantial
396 parental frequency differentials – 0.6), rather than a random subset that might contain many

397 uninformative loci, as suggested in other studies (Knief et al. 2019; Chaturverdi et al. 2020;
398 Del-Rio et al. 2021; Walsh et al. 2023) and following recommendations in Martin et al.
399 (2021). In our *BGC* analysis, positive α values indicate excess *P. glauca* ancestry compared
400 to the genome-wide average, while negative α values indicate excess *P. engelmannii*
401 ancestry. On the other hand, positive β values indicate a steeper cline and selection against
402 introgression resulting in a faster rate of allelic-frequency transition along the hybrid index,
403 whereas negative β values pinpoint loci that are introgressed more than expected, indicating
404 some sort of hybrid vigour (i.e. complementation of deleterious recessive).

405 In the elevational dataset, pre-processing resulted in a set of 3,256 SNPs for genomic clines
406 analysis, which in turn led to the identification of 300 SNPs exhibiting excess ancestry (148
407 positive α and 152 negative α SNPs). These loci represent SNPs with introgression patterns
408 different from the genome-wide average, with a 95% equal-tail probability interval (ETPI) for
409 the posterior probability distribution of α that did not overlap with zero. No SNP showing
410 unusual rates of allelic-frequency transition along the hybrid index was detected in the
411 elevational dataset genomic clines analysis (β excess) (supplementary figs. S18 and S19,
412 Supplementary Material online).

413 In the latitudinal dataset, pre-processing resulted in a set of 10,154 SNPs for genomic clines
414 analysis, which in turn led to the identification of 3,845 SNPs with a 95% ETPI for the
415 posterior probability of α that did not overlap 0 (1,906 positive α and 1,939 negative α
416 SNPs), indicating excess ancestry (supplementary figs. S19 and S20, Supplementary
417 Material online). Additionally, the latitudinal dataset analysis also pinpointed 120 SNPs
418 exhibiting unusual rates of allele frequency transitions in the hybrid population (52 negative β
419 and 68 positive β SNPs) (supplementary fig. S20, Supplementary Material online). Similarly
420 to SNPs exhibiting excess ancestry (α excess SNPs), these SNPs reported a 95% ETPI for
421 the posterior probability distribution of β that did not overlap with zero (β excess). Among the
422 SNPs that reported α or β excess in the latitudinal dataset analysis, 76 were in common,
423 indicating loci exhibiting both excess ancestry as well as unusual rates of allelic-frequency

transitions along the hybrid index gradient (supplementary fig. S20, Supplementary Material online).

Furthermore, we performed two additional *BGC* analyses on subsets of the latitudinal dataset, referred to as Subset 1 and Subset 2, by downsampling the data to a similar number of individuals and parental population composition as in the elevational dataset (supplementary fig. S21, Supplementary Material online). This was done to determine whether the high number of α-excess SNPs observed in the latitudinal dataset, relative to the total number of SNPs tested, was a statistical artifact of their differing sample sizes, or indicated a stronger biological signal in the latitudinal transects compared to the elevational transects. Both downsampled analyses revealed a pattern closely resembling that of the full dataset analysis, identifying 3,258 and 3,268 α-excess SNPs, respectively (supplementary fig. S22, Supplementary Material online), suggesting a stronger biological signal in the latitudinal dataset. Furthermore, the Pearson's correlation between α values from the full latitudinal dataset analysis and those from the subset analyses was very high, at 0.91 and 0.90 (p-value < 2.2e-16). For comparison, α correlation between the two subset analyses was even higher, at 0.97 (p-value < 2.2e-16).

Comparing across datasets, the elevational and latitudinal dataset clines analyses shared a total of 982 SNPs. These showed significant correlation of α values between the elevational and latitudinal dataset analyses (Pearson's correlation: 0.32, p-value < 2.2e-16; Kendall's tau: 0.21, p-value < 2.2e-16) (Fig. 4d). In more detail, within the 982 shared SNPs, 438 exhibited excess ancestry in the latitudinal dataset analysis and 83 in the elevational – between these, 53 reported excess ancestry across both analyses, with 49 (92%) exhibiting excess ancestry in the same direction of introgression.

Overall, *BGC* α values were generally in agreement with the p-value obtained with *introgress* (between *introgress* p-values and *BGC* absolute α values) for both datasets (supplementary fig. S23, Supplementary Material online).

450 By visually inspecting Manhattan plots derived from our genomic clines analyses, it is
451 apparent that some of the strongest GEA peaks comprising our top candidate for
452 environmental association show asymmetric introgression from one parental species to the
453 other (excess ancestry). This could indicate local adaptation associated with the genomic
454 region's ancestry, even though we cannot exclude stable genetic differentiation (i.e. genetic
455 drift) (Fig. 4a and c; supplementary fig. S24, Supplementary Material online). SNPs with
456 these asymmetric patterns tend to be organized into large blocks that all tend to be
457 introgressed in the same direction (Fig. 4a and c; supplementary fig. S24, Supplementary
458 Material online).

459

460 **Candidate Genes for Local Adaptation**

461 The elevational dataset had coverage of 2,901 genes while the latitudinal dataset covered
462 6,901 genes, with 2,353 of these genes included in both datasets. In the elevational dataset
463 analysis, we found 138 genes harbouring at least one of the 1% lowest GEA p-values SNPs
464 and 48 genes harbouring at least one of the SNPs that reported excess ancestry in the
465 genomic clines analysis (30 candidate genes in common), for a total of 156 unique candidate
466 adaptive genes (supplementary fig. S25, Supplementary Material online). In the latitudinal
467 dataset analysis, we found 297 genes harbouring the 1% lowest GEA p-values SNPs and
468 466 genes harbouring the SNPs that reported excess ancestry in the genomic clines
469 analysis (173 candidate genes in common), for a total of 590 unique candidate adaptive
470 genes (supplementary fig. S25, Supplementary Material online). Between these two dataset-
471 specific sets of candidate genes, 131 genes were in common, out of 615 unique candidate
472 adaptive genes identified. In more detail, 108 genes were identified as candidates by both
473 GEAs (top 1%), 29 genes by both clines analyses, 75 genes between the elevational dataset
474 GEA and the latitudinal dataset clines analysis, and 35 between the elevational dataset
475 clines analysis and the latitudinal dataset GEA (supplementary fig. S25, Supplementary
476 Material online). In particular, 19 genes were identified as adaptive candidates by all four

477 analyses (both GEAs and both clines analyses) (supplementary Table S1 and
478 supplementary fig. S25, Supplementary Material online). Although this set requires further
479 study and validation, it included some protein families with key roles in plant growth and
480 environmental stress response, according to the available annotation (Warrant et al. 2015;
481 Gagalova et al. 2022). For instance, cytochrome P450, a superfamily of enzymes linked to
482 both biotic and abiotic stress responses as well as developmental processes in plants (Xu et
483 al. 2015), which has also been recently identified as a target of repeated global selection in
484 multiple plant species (Nocchi et al. 2024). MYB transcription factors, which are an
485 extensively studied family of plant proteins involved in development, metabolism and
486 responses to biotic and abiotic stresses (Biswas et al. 2023; Roy 2016). The HhH-GPD DNA
487 glycosylase superfamily, which includes proteins involved in DNA repair caused by oxidative
488 stress (Nota et al. 2015). DEAD/DEAH box RNA helicases, which mediate plant
489 development and abiotic stress tolerance through modulation of pre-rRNA processing (Liu
490 and Imai 2018). And finally, a family of UV-B-induced proteins, though not yet fully
491 characterized, was also present within this set of 19 genes.

492 Additionally, we identified three genes that appear to be involved in adaptation to elevational
493 but not latitudinal, environmental pressures (supplementary fig. S25, Supplementary Material
494 online). These genes belonged to the no apical meristem (NAM) and F-box protein families.
495 NAM proteins are involved in responses to various abiotic factors and play key roles in plant
496 development, particularly in leaf formation, shaping morphology, patterning, and size, which
497 in turn influence critical processes such as transpiration, light absorption, and thermal
498 regulation (Hasson et al. 2010; Hu et al. 2010; Cheng et al. 2012; Wang et al. 2022). These
499 proteins have also been previously identified as adaptive candidates in common garden
500 studies of local adaptation in interior spruce in British Columbia (De La Torre et al. 2014). F-
501 box proteins are similarly involved in a wide range of responses to abiotic factors, including
502 drought, temperature, and in particular, UV radiation (Lechner et al. 2006; Zhang et al. 2013,
503 2014; An et al. 2016; Wang et al. 2024). These types of abiotic factors tend to vary more

504 abruptly with elevation than with latitude, as environmental transitions along latitude are
505 mostly related to photoperiod and seasonality and generally shift more gradually. While this
506 interpretation is consistent with known functional roles, we note that further validation is
507 needed to confirm the involvement of these genes in adaptation specific to elevational
508 gradients.

509 On the other hand, we identified 103 genes (supplementary fig. S25, Supplementary
510 Material online) that appear to be involved in adaptation to latitudinal, but not elevational,
511 factors. While exploring each candidate goes beyond the scope of this work and this set
512 requires further validation, it is worth highlighting the presence of several proteins involved in
513 photoperiodic flowering regulation. These included the phosphatidylethanolamine-binding
514 protein (PEBP) family (Faure et al. 2007), COL genes (Chen et al. 2021), PP2C
515 phosphatases (Zeng et al. 2022), TCP transcription factors (Zhang et al. 2024), AP2/ERF
516 transcription factors (Matías-Hernández et al. 2014), SPL transcription factors (Jung et al.
517 2011), 14-3-3 proteins (Paul et al. 2008) and the trehalose-phosphatase family (Wahl et al.
518 2013).

519

520 **Comparison with adaptive loci identified in similar studies of local adaptation in**
521 **conifers**

522 The study by Hornoy et al. (2015) identified 131 *P. glauca* genes with signals of selection
523 along temperature and precipitation gradients in eastern Canada, of which 108 were
524 included in our tested gene set. Between these 108 genes and the relaxed set of 615 genes
525 identified as candidates in either of our analyses, 7 were shared (supplementary Table S2,
526 Supplementary Material online). This study tested about a quarter of the transcriptome
527 (7,819), but since it is not clear exactly which genes were tested, a hypergeometric
528 expectation cannot be calculated. Nonetheless, it is unlikely that the identified overlap (7
529 genes) represents a strong enrichment against random expectation. The set of seven

common candidate genes included some notable protein families, such as the AP2/ERF transcription factor family, which comprises essential regulators in many biological processes like plant growth, fruit/flower development, photoperiodic flowering and biotic and abiotic stress responses (Matías-Hernández et al. 2014; Jiang et al. 2022; Jiang et al. 2024). This family has also been identified as a target of repeated global selection across several plant species in a recent comparative genomics study (Nocchi et al. 2024). RieskeFeS proteins which are involved in light capture have been shown to improve the efficiency of photosynthesis and to stimulate growth in low light conditions (Simkin et al. 2017). And the PAP2 superfamily, which includes proteins that when overexpressed have been shown to increase photosynthetic rate and enhance growth (Cai et al. 2022).

The study on red spruce (*P. rubens* Sarg.) in north-eastern North America by Capblancq et al. (2022) identified 125 genes with signals of local adaptation along climatic gradients, of which 76 had orthology with *P. glauca* according to our orthology assignment. Between these 76 genes and the 615 genes identified in our study (of which 541 with orthology), 7 were in common (supplementary Table S2, Supplementary Material online). In total our analysis and their analysis tested 4,897 common orthogroups. Although the observed overlap of 7 genes does not significantly deviate from the statistical expectation based on the hypergeometric distribution (expected overlap = 8.39, observed overlap = 7, p-value = 0.75), this set of common candidates included some notable proteins, such as a longevity assurance gene (LAG1) which helps plants resist toxins and survive ER stress (Brandwagt et al. 2000), and a flowering time control gene (FPA). FPA has been shown to regulate flowering via a temperature-driven but day-length independent mechanism, suggesting its potential role in the phenological differentiation of populations experiencing identical photoperiod regimes (i.e. same latitude, different elevation) (Schomburg et al. 2001; Capblancq et al. 2022).

The study by Yeaman et al. (2016) on convergent evolution identified 258 lodgepole pine candidate genes for local adaptation along geographic and environmental gradients in north-

western North America, of which 136 had orthology relationships that could be identified with spruce. Between these 136 genes and the 615 genes identified in our study (of which 489 had orthology with pine), 12 were shared (supplementary Table S2, Supplementary Material online). In total these two analyses tested 4,220 common orthogroups. This set of 12 genes, despite not significantly differing from the number of expected overlap based on the hypergeometric distribution (expected overlap = 15.75, observed = 12, p-value = 0.87), still included some notable proteins, such as TCP transcription factors which mediate plant growth and development (Yu et al. 2022), AP2/ERF transcription factors (also a candidate protein family in Hornoy et al. 2015), which are involved in growth, flowering regulation and defence responses to a variety of stresses (Matías-Hernández et al. 2014; Jiang et al. 2022; Jiang et al. 2024), and four zinc finger proteins (ZFPs). These latter proteins constitute the largest E3 ubiquitin ligase family and play important roles in plant growth, development, and response to abiotic stresses such as drought, salt, temperature, reactive oxygen species and harmful metals (Han et al. 2022). Consistent with this, the study by Depardieu et al. (2021) on drought adaptation in *P. glauca* in Quebec (Canada) found that ZFPs were among the most represented gene families in their set of adaptive candidates. More specifically, Depardieu et al. (2021) identified eight high-confidence genes associated with drought tolerance in white spruce populations, including GQ03707_G19 (DB47_00099276), a zinc finger transcription factor that was also detected in our study and in the lodgepole pine analysis in the study by Yeaman et al. (2016). This gene was correlated with the summer soil moisture index in their GEA, was linked to annual fluctuations in wood density in their genotype–phenotype association (GPA), and was also shown to be upregulated under drought conditions in their gene expression analysis (Depardieu et al. 2021). Furthermore, a ZFP was also present in the set of 19 candidate genes identified by all our analyses, and E3 ubiquitin ligases were recently revealed as key drivers of global adaptation across several plant species in Nocchi et al. (2024).

584 ## Discussion

585 Here, we assessed local adaptation, hybridization, and introgression patterns in two regions
586 of the Western Canada interior spruce hybrid zone: a latitudinal gradient spanning >1000 km
587 of gradual environmental change and a narrower elevational contact zone spanning
588 ~100km, including many regions with dramatic changes in environment over short spatial
589 scales of < 10km. Our analyses revealed that patterns in the genomic basis of introgression
590 are broadly similar between the elevational and latitudinal transects. We found strong
591 repeatability in our F_{ST} and GEA analyses, with high correlation of the resulting F_{ST} and GEA
592 p-values and a visually concordant distribution of peaks and valleys in the resulting
593 Manhattan plots (Figs. 3 and 4; supplementary figs. S11 and S14, Supplementary Material
594 online). Additionally, directional introgression patterns also appeared broadly consistent
595 between the elevational and latitudinal hybrid zones (Fig. 4; supplementary fig. S24,
596 Supplementary Material online).

597 Repeatability of patterns in the genome across replicates only provides strong evidence of
598 natural selection if the replicates represent independent bouts of evolution. Here, because of
599 the shared parental lineages that gave rise to these different zones of contact and
600 hybridization, the latitudinal and elevational transects are non-independent so the similarity
601 in F_{ST} and GEA patterns alone cannot be taken as strong evidence of natural selection.
602 Genome-wide variation in the extent of lineage sorting within parental species prior to
603 secondary contact could explain a substantial proportion of the observed similarity in F_{ST},
604 and similar GEA patterns would then also occur due to covariation with the environment
605 across the hybrid zone. That said, it is likely that patterns in many regions with high F_{ST} and
606 strong GEA signal were driven by natural selection, as there is considerable evidence of
607 local adaptation based on reciprocal transplant experiments (Jaramillo-Correa et al. 2001;
608 De La Torre et al. 2013; Liepe et al. 2016), but the proportion cannot be quantified based on
609 the available evidence. On the other hand, the repeated patterns of directional introgression
610 represent stronger evidence of selection, as such repeated departures from genome-wide

patterns would be unlikely to occur by chance due to drift and migration. Future work using genome-wide association studies of climate stress phenotypes in common gardens could further improve our understanding of which areas of the genome are under strong divergent selection. Our analysis also highlights that correcting for population structure in GEAs can be counterproductive, particularly when there is strong collinearity between the environmental gradient and the axis of introgression (supplementary fig. S17, Supplementary Material online). Methods that include structure correction may fail to detect true adaptive loci in such cases, as previously suggested (Lotterhos 2019; Booker et al. 2023).

In our analysis of the genomic distribution of regions of high differentiation between species, chromosomal autocorrelation of the resulting F_{ST} and GEA p-values shows that differences in the length of chromosomal tracts of introgression between latitudinal and elevational datasets are broadly consistent with expectations based on the theory of hybridization (Barton 1983; Sedghifar et al. 2016). In the elevational transects, there is less opportunity for recombination across the hybrid zone, because of the abrupt transition across shorter spatial distances (Shchur et al. 2020), which results in larger blocks of ancestry around the loci with strong patterns of differentiation and association with environment (supplementary figs S12 and S16, Supplementary Material online). Given that common garden experiments show substantial evidence for local adaptation between the parental species, many of these regions are likely shaped by strong divergent selection, which would further contribute to these patterns (Shchur et al. 2020). The power of this analysis was limited by the fragmented state of the spruce genome assembly and limited overlap in the genomic regions targeted by the different sequence capture panels used for the two datasets, so further exploration of this hypothesis is warranted.

The candidate adaptive loci identified included several genes previously reported for their central role in plant growth and environmental adaptation. Some of these genes were also recently identified in similar studies of local adaptation in other North American conifers (Hornoy et al. 2015; Yeaman et al. 2016; Depardieu et al. 2021; Capblancq et al. 2022).

638 Additionally, differences in candidate loci between the elevational and latitudinal transects
639 revealed several genes potentially involved in the response to environmental factors specific
640 to latitude such as photoperiod and seasonality, as well as genes potentially involved in
641 regulating factors more specific to elevational changes, such as oxygen levels, UV radiation,
642 temperature, snow and precipitation levels. However, we note that our candidates for local
643 adaptation require further validation, and it is likely that this set includes false positives due
644 to the difficulty in completely disentangling environmentally driven variation from neutral
645 population structure effects. One way we tried to address this was through the analysis of
646 genomic clines, as directional introgression replicated across hybrid zones characterized by
647 different environmental contrasts would not be expected under a neutral drift model
648 (Rieseberg et al. 1999; Buerklen and Lexer 2008; Gompert and Buerkle 2009; Bierne et al.
649 2013; Simon et al. 2020; Westram et al. 2021; Langdon et al. 2022). This is what we
650 observed: directional introgression occurred largely in the same genomic regions and often
651 in the same direction between the elevational and latitudinal transects, as reflected by the
652 significant correlation of *BGC* alpha scores and overall highly concordant genome-wide
653 patterns of directional introgression (Fig. 4; supplementary fig. S24, Supplementary Material
654 online). This supports the idea that the 19 adaptive loci retrieved by all four analyses form a
655 robust set of adaptive loci (supplementary Table S1 and supplementary fig. S25,
656 Supplementary Material online), unlikely to be generated by drift.

657 Finally, our results also echo previous work suggesting that the interior spruce hybrid zone is
658 maintained by a bounded hybrid superiority model, where hybrids exhibit higher fitness than
659 parental species in intermediate environments (De La Torre et al. 2013, 2014). In our case,
660 this may be represented by south-facing high-altitude slopes of the elevational dataset,
661 which combine features of the warmer, drier valleys preferred by *P. glauca* and the colder,
662 wetter and later-melting north-facing slopes favoured by *P. engelmannii*. To suggest this,
663 south-facing slopes reported a higher frequency of balanced hybrids (hybrids close to 50:50
664 ratio), a balanced distribution of *P. glauca* dominant and *P. engelmannii* dominant hybrids as

665 well as lower frequencies of putative pure *P. glauca* individuals compared to valleys, and
666 putative pure *P. engelmannii* individuals compared to north-facing slopes (Fig. 2;
667 supplementary fig. S9, Supplementary Material online). The intermediate habitats provided
668 by south-facing slopes may allow hybrids to outperform both parental species by combining
669 traits such as faster growth with earlier bud setting and greater tolerance to snow load (De
670 La Torre et al. 2013, 2014).

671

672 Materials and Methods

673 Super-scaffold assembly construction

674 We built a super-scaffold assembly of the WS77111v2 white spruce (*P. glauca*) genome to
675 use in subsequent analyses, by conservatively anchoring scaffolds to the latest publicly
676 available *P. glauca* linkage map including 14,727 cDNAs (Gagalova et al. 2022). For this
677 process, we retrieved the sequences of an exhaustive set of 27,720 cDNAs, stemming from
678 the catalogue of spruce expressed genes assembled by Rigault et al. (2011), which we
679 filtered to retain only the sequences of the 14,727 cDNAs present in the *P. glauca* linkage
680 map (Gagalova et al. 2022). We then used *GMAP* (Wu et al. 2016) to align the retained
681 cDNAs sequences to the scaffolds of the WS77111v2 white spruce reference genome, and
682 the output was processed according to the following criteria: mapped cDNAs with coverage
683 above 50% and sequence identity above 90% were retained, while anything else was
684 discarded to ensure retention of only high confidence hits. Then, for each cDNA that satisfied
685 these mapping criteria, we extracted its best scaffold match based on the highest sequence
686 identity. In case a cDNA best mapping scaffold could not be determined as it mapped equally
687 well to multiple locations in the genome, it was discarded. This conservative approach
688 ensured the removal of difficult to resolve multi-mapped cDNAs. Finally, the scaffolds with
689 mapped cDNAs were ordered along the 12 spruce linkage groups according to the cDNAs
690 centimorgan (cM) positions in the linkage map.

691

692 **Sampling and sequencing**

693 We collected cambium tissue material from 331 mature spruce individuals, including *P.*
694 *glauca*, *P. engelmanni* and their hybrids, from seven transects along the Trans-Canada
695 highway going through Canmore, Banff and Yoho in the summer of 2022 (Fig. 1c, elevational
696 transects). Each transect included three distinct populations subjected to contrasting
697 environmental pressures: a north-facing high-altitude population, a south-facing high-altitude
698 population, and a valley population (Fig. 1c). Within each population, we sampled between
699 14 and 16 individuals (supplementary fig. S1, Supplementary Material online). Finally,
700 cambium tissue samples were sent to Rapid Genomics LLC (Gainesville, Florida) for DNA
701 extraction, library preparation, and sequencing using a 10 Mb target approach.

702 Additionally, we assembled a second dataset by downloading raw exome capture data for
703 575 mature spruce individuals from across 249 populations distributed throughout western
704 Canada (Fig. 1a, latitudinal transects), derived from previous studies (Suren et al. 2016;
705 Yeaman et al. 2016). This data is publicly available in the Short Read Archive (SRP071805;
706 PRJNA251573).

707

708 **SNP calling**

709 We applied a uniform SNP calling pipeline to both datasets for consistency. First, *fastp*
710 (Chen et al. 2018) was used to remove sequencing adapters and for trimming raw paired
711 end FASTQ files, using default settings. Processed reads were then mapped to the latest
712 publicly available *P. glauca* reference genome (WS77111v2,
713 https://www.bcgsc.ca/downloads/btl/Spruce/) (Warren et al. 2015) using *BWA-mem* (Li and
714 Durbin 2009). Following mapping, we used *samtools* (Li et al. 2009) to convert the alignment
715 files generated from SAM to sorted BAM formats and discarded any alignment with mapping
716 quality below 10. *Picard* tools (available at https://broadinstitute.github.io/picard/) was then

717 used to remove potential PCR duplicates and to assign read groups, followed by indel
718 realignment with *GATK 3.8* (DePristo et al. 2011). Finally, SNP calling was performed using
719 *BCFtools mpileup* and *BCFtools call* (Danecek et al. 2021), based only on alignments with a
720 minimum mapping quality of 5 (−q 5). We filtered raw VCF files with *VCFtools* (Danecek et
721 al. 2011) to retain only biallelic SNPs with the following characteristics: genotyped in at least
722 70% of the individuals, quality above 30 (--minQ 30), individuals genotype quality above 20
723 (--minGQ 20), minimum read depth above 5 (--minDP 5) and minor allele frequency greater
724 than 0.05. Additionally, we filtered the VCF files of both datasets (latitudinal and elevational)
725 to retain only the SNPs located on the scaffolds that we were able to anchor to the available
726 linkage map (Gagalova et al. 2022) in our super-scaffold assembly construction.

727

728 **Population structure assessment**

729 We used *fastSTRUCTURE* (Raj et al. 2014) to assess population structure in both datasets,
730 using a simple prior and testing Ks ranging from 1 to 10, and then selected the K model that
731 maximizes the log-marginal likelihood lower bound (LLBO) of the data (K*ε). To determine
732 which species each *fastSTRUCTURE* cluster represented, we added whole genome
733 sequencing data of two pure species individuals cultivated at the arboretum of the Horsholm
734 University of Copenhagen (Denmark) and repeated the *fastSTRUCTURE* inference. This
735 data was retrieved from ENA under accessions SAMEA51442168 (*P. glauca*) and
736 ERX1862345 (*P. engelmannii*). Furthermore, we classified individuals as either pure *P.*
737 *glauca*, pure *P. engelmannii* or hybrid according to the value of the admixture coefficient (Q)
738 computed with *fastSTRUCTURE* at the chosen model complexity (K=2): $q \geq 0.95$ for either
739 pure species, while individuals with $0.05 < q < 0.95$ were classified as hybrids. Population
740 structure was further assessed using principal component analysis (PCA), implemented in
741 *Plink* version 2.0 (Chang et al. 2015). Marker-based realized genomic relatedness was
742 computed between all possible pairs of individuals according to the formula by VanRaden
743 (2008), implemented in the kin function of the R package *synbreed* (Wimmer et al. 2012).

744

745 **Population differentiation**

746 Weir and Cockerham F_{ST} (Weir and Cockerham 1984) was calculated on a per-SNP basis
747 between the 21 populations of the elevational dataset and the 249 populations of the
748 latitudinal dataset using the *VCFtools* function *weir-fst-pop* (Danecek et al. 2011).
749 Additionally, F_{ST} between *P. glauca* and *P. engelmannii* was also calculated within each
750 dataset, selecting pure species individuals based on the admixture coefficient computed with
751 *fastSTRUCTURE* ($q \geq 0.95$). Moran's I indices of spatial autocorrelation based on F_{ST} scores
752 were computed using the R package *moranfast* (https://github.com/mcooper/moranfast),
753 using SNP position along both scaffold and linkage group. To measure clustering of F_{ST}
754 values along the genome, we calculated Moran's I using SNP index positions, assigning
755 each SNP a sequential number (1, 2, 3, ...) based on its genomic order in our super-scaffold
756 assembly. This approach treats SNPs as equally spaced and captures how long similar
757 values persist along scaffolds and, more broadly, along linkage groups. It is intended to
758 reflect the blockiness of the signal, which may be related to haplotype structure. We used
759 ordinal position of SNPs rather than genomic positions because calculation of Moran's I
760 using genomic position would be lower if a given set of SNPs were spread across long
761 physical distances. As it is expected that SNP density would be lower in regions with slow
762 LD decay and longer LD tracts, using genomic distances would erroneously bias the
763 measurement downwards in these regions. Ideally, this analysis would use genetic distance,
764 but we do not have suitable genomic resources to enable this.

765

766 **Genome environment association analysis**

767 We downloaded elevation data at 90-meter resolution from the NASA Shuttle Radar
768 Topography Mission (SRTM) (https://doi.org/10.5069/G9445JDF) and used it to compute
769 slope and aspect using the *terrain* function of the *raster* R package (Hijmans and Etten

2012). Additionally, we downloaded seven solar variables from the Global Solar Atlas (https://globalsolaratlas.info): photovoltaic electricity output (PVOUT), global horizontal irradiation (GHI), diffuse horizontal irradiation (DIF), direct normal irradiation (DNI), optimum tilt-angle (OPTA), air temperature at 2 meters (TEMP) and global tilted irradiation (GTI).

After extracting measurements for this set of 11 variables for all our samples, we selected environmental variables with Pearson's correlation coefficient < 0.7 in order to avoid overfitting. The following five variables were retained: slope, aspect, DIF, PVOUT and TEMP. To describe the selected solar variables, DIF measures the scattered sunlight that reaches a horizontal surface, excluding direct solar beams. It serves as an indicator of indirect light availability, which can influence plant adaptation in shaded or cloudy environments. In contrast, PVOUT represents the potential electricity generated by solar panels and acts as a proxy for total solar irradiance, integrating both direct and indirect light. As a result, PVOUT provides a measure of the total environmental solar energy available at a given location.

Moreover, we generated high-resolution (30 meter) snow-cover data for the region covered by the elevational dataset using the R package *SnowWarp* (Berman et al. 2018; Laurin et al. 2022), which fuses 30-meter Landsat optical imagery with 500-meter MODIS (Moderate Resolution Imaging Spectroradiometer) snow cover data using the Google Earth Engine platform (Gorelick et al. 2017). Daily fractional snow-covered area (fSCA) was computed for the time period spanning 2000 to 2021 and was used to derive the yearly date of snowmelt for each 30-meter pixel, which was calculated as the first date in the second half of a snow year where at least 15 of the previous 30 days were snow-free (Berman et al. 2019). Finally, the mean snowmelt date was obtained by averaging snowmelt dates from 2000 to 2021 for each of the 21 populations in the elevational dataset. For each population, before calculating the yearly average snowmelt date, we applied a spatial smoothing step by averaging the snowmelt dates of each pixel together with its neighbouring pixels, rather than using the exact value of each population location pixel alone. This approach aimed to capture a more representative local signal around each population. It was not feasible to generate this high-

797 resolution data for the vast area covered by latitudinal dataset, therefore we only included
798 snowmelt day for the elevational dataset GEA, as we deemed it particularly well suited to
799 discriminate between north-facing slopes, south-facing slopes and valleys (supplementary
800 fig. S2, Supplementary Material online).

801 We then used redundancy analysis (RDA) (Capblancq and Forester 2021) in the R package
802 *vegan* to identify the linear combination of the retained environmental variables that explains
803 the greatest variation in the hybrid index. This was performed independently for the
804 elevational (six environmental variables including snow melt) and latitudinal dataset (five
805 environmental variables). Finally, we used R to perform the genotype environment
806 association analyses. These involved a likelihood-ratio test for each SNP, comparing a null
807 intercept-only model to a predictor model which used the first axis of variation (RDA1) from
808 RDA to predict SNP frequency. In a likelihood-ratio test, a low p-value is obtained when a
809 significantly better fit is obtained with the predictor model and it reflects the importance of the
810 predictor: the smaller the p-value, the more important the predictor.

811 Following the same approach used in the F_{ST} analysis, we computed Moran's I indexes of
812 spatial autocorrelation based on GEA p-values using the R package *moranfast*
813 (https://github.com/mcooper/moranfast). To provide a comparison, we repeated the log-
814 likelihood ratio test-based GEA analysis, this time incorporating structure correction using
815 the first principal component (PC1) from our PCA. In this analysis, we compared a null
816 model, where SNP frequency was predicted by PC1 alone, to a predictor model that
817 included both PC1 and the first axis of variation from RDA (RDA1 * PC1) to predict SNP
818 frequency.

819

820 **Genomic clines analysis**

821 We performed a genomic clines analysis to identify SNP loci exhibiting excess ancestry from
822 either parental species relative to the average genomic background, as well as to identify

823 SNP loci with unusual rates of allelic-frequency transitions along the hybrid index. For this
824 analysis we used a Bayesian genomic clines model implemented in *BGC* (Gompert and
825 Buerkle 2012) and a logistic regression-based model implemented in *introgress* (Gompert
826 and Buerkle 2010). In *introgress*, we used both permutation and parametric tests to compute
827 p-values. We restricted each analysis to SNP loci with a frequency differential between pure
828 *P. glauca* and pure *P. engelmannii* of at least 0.6. Parental pure species and hybrid
829 populations were defined for both datasets according to the admixture coefficient computed
830 with *fastSTRUCTURE* ($Q \geq 0.95$). This resulted in the following setups: 82 pure *P.*
831 *engelmannii*, 163 pure *P. glauca*, and 330 admixed individuals in the latitudinal dataset, and
832 99 pure *P. engelmannii*, 40 pure *P. glauca*, and 192 admixed individuals in the elevational
833 dataset. Furthermore, we conducted two additional *BGC* analyses on the latitudinal dataset,
834 using the same SNPs but downsampling the data to obtain a similar number of individuals
835 and parental population composition as in the elevational dataset, to allow for better
836 comparison. Subset 1 included 70 pure *P. engelmannii*, 86 pure *P. glauca*, and 200 admixed
837 individuals, while Subset 2 included 70 pure *P. engelmannii*, 45 pure *P. glauca*, and 194
838 admixed individuals. In our subsampling strategy, we prioritized removing individuals from
839 the latitudinal dataset which were located near the area sampled in the elevational dataset,
840 as well as populations from western British Columbia (supplementary fig. S21,
841 Supplementary Material online), to avoid possible signals of admixture with Sitka spruce
842 (Hamilton et al. 2012). In *BGC* we used 100,000 MCMC generations with 50,000 burn-in and
843 retained every 24th iteration. After assessing convergence, SNPs showing excess ancestry
844 or unusual rates of allelic-frequency transitions along the hybrid population were identified as
845 those for which the 99% equal-tail probability interval (ETPI) for the posterior probability
846 distribution of the clines parameters α or β did not overlap zero. In *introgress*, we used
847 default settings and ran both the parametric and permutation tests (Gompert and Buerkle
848 2010). Both tools were run within the *ClineHelpR* R package (Martin et al. 2021).

849

850 **Identification of candidate genes and overlap with other studies**

851 We used bedtools (Quinlan and Hall 2010) to identify the genes intersected by our top
852 candidates SNPs from the GEA and genomic clines analyses. We expanded the gene
853 sequences by adding 5000 bp flanks on either side, to include possible regulatory elements.
854 We used the gene models provided for the WS77111v2 *P. glauca* reference genome
855 (https://www.bcgsc.ca/downloads/btl/Spruce/). We used *Orthofinder2* (Emms and Kelly
856 2019) on each species proteome to identify orthology between white and red spruce, and
857 between white spruce and lodgepole pine. For white spruce (*P. glauca*), we used the
858 proteome for WS77111v2, available at
859 https://www.bcgsc.ca/downloads/btl/Spruce/Pglauca_WS77111/genome_annotation/WS7711
860 1v2/. As red spruce was mapped to Norway spruce (*P. abies*) in the study by Capblancq et
861 al. (2022), we used the proteome available for *P. abies* version 1 reference genome
862 (https://treegenesdb.org/org/Picea-abies). For lodgepole pine (*Pinus contorta*), we first
863 retrieved the entire transcriptome analysed in Yeaman et al. (2016) and used TransDecoder
864 (https://github.com/TransDecoder/TransDecoder) to identify the longest open reading frames
865 for each transcript and then translate them into protein sequences. Finally, for each species
866 proteome we selected a single primary transcript per gene according to the longest isoform,
867 and then *Orthofinder2* was run using default settings and independently for white spruce
868 versus Norway spruce and white spruce versus lodgepole pine.

869

870 Acknowledgements

871 We would like to thank Sally Aitken for many conversations that helped inspire this work, Ian
872 MacLachlan, Jason Mogilefsky, Vashti Dunham, and Tim McCready for help obtaining
873 funding and overall planning of the project, and staff at the Canadian and Alberta parks
874 systems for help with securing permits. Thanks also to Michael Williamson for help
875 brainstorming the methods, to members of the Yeaman lab for feedback, and to Janek's

brother, Grant Scully, and Sai Yeaman for help with field work. This project was funded by an NSERC Alliance and FRIAA funds contributed by Weyerhauser and West Fraser Timber, with computational support from the Digital Research Alliance of Canada.

Competing Interests

The authors declare no competing interests.

Data Availability

Sequence data are deposited in the Short Read Archive under BioProject IDs PRJNA251573 (latitudinal dataset) and PRJNA1294281 (elevational dataset).

References

1. Aguillon SM, Dodge TO, Preising GA, Schumer M. 2022. Introgression. *Current Biology* 32.
2. An J-P, Li R, Qu F-J, You C-X, Wang X-F, Hao Y-J. 2016. Apple F-box protein mdmax2 regulates plant photomorphogenesis and stress response. *Frontiers in Plant Science* 7.
3. Anderson E. 1953. Introgressive hybridization. *Biological Reviews* 28:280–307.
4. Baack EJ, Rieseberg LH. 2007. A genomic view of introgression and hybrid speciation. *Current Opinion in Genetics & Development* 17:513–518.
5. Barton NH, Hewitt GM. 1985. Analysis of hybrid zones. *Annual Review of Ecology and Systematics* 16:113–148.
6. Barton NH. 1983. Multilocus Clines. *Evolution* 37:454.
7. Becker M, Gruenheit N, Steel M, Voelckel C, Deusch O, Heenan PB, McLenachan PA, Kardailsky O, Leigh JW, Lockhart PJ. 2013. Hybridization may facilitate in situ survival of endemic species through periods of climate change. *Nature Climate Change* 3:1039–1043.
8. Berman EE, Bolton DK, Coops NC, Mityok ZK, Stenhouse GB, Moore RD (Dan). 2018. Daily estimates of landsat fractional snow cover driven by Modis and dynamic time-warping. *Remote Sensing of Environment* 216:635–646.
9. Berman EE, Coops NC, Kearney SP, Stenhouse GB. 2019. Grizzly Bear Response to fine spatial and temporal scale spring snow cover in western Alberta. *Plos One* 14.
10. Bierne N, Gagnaire P-A, David P. 2013. The geography of introgression in a patchy environment and the Thorn in the side of ecological speciation. *Current Zoology* 59:72–86.
11. Biswas D, Gain H, Mandal A. 2023. MYB transcription factor: A new weapon for biotic stress tolerance in plants. *Plant Stress* 10:100252.

12. Booker T, Yeaman S, Whitlock M. 2023. The WZA: A window-based method for characterizing genotype-environment association. *Mol. Ecol. Resour.* 24(2):e13768.
13. Brandwagt BF, Mesbah LA, Takken FL, Laurent PL, Kneppers TJ, Hille J, Nijkamp HJ. 2000. A longevity assurance gene homolog of tomato mediates resistance to *Alternaria alternata f. sp.. lycopersici* toxins and fumonisin B_1. *Proceedings of the National Academy of Sciences* 97:4961–4966.
14. Brennan A, Woodward G, Seehausen OL, Muñoz-Fuentes V, Moritz C, Guelmami A, Abbott R, Edelaar P. 2014. Hybridization due to changing species distributions: Adding problems or solutions to conservation of biodiversity during global change? *Evolutionary ecology research* 16:475-491.
15. Buerkle CA, Lexer C. 2008. Admixture as the basis for genetic mapping. *Trends in Ecology & Evolution* 23:686–694.
16. Cai Y, Zhai Z, Blanford J, Liu H, Shi H, Schwender J, Xu C, Shanklin J. 2022. Purple acid phosphatase2 stimulates a futile cycle of lipid synthesis and degradation, and mitigates the negative growth effects of triacylglycerol accumulation in vegetative tissues. *New Phytologist* 236:1128–1139.
17. Capblancq T, Forester BR. 2021. Redundancy analysis: A Swiss Army Knife for Landscape Genomics. *Methods in Ecology and Evolution* 12:2298–2309.
18. Capblancq T, Lachmuth S, Fitzpatrick MC, Keller SR. 2022. From common gardens to candidate genes: Exploring local adaptation to climate in Red Spruce. *New Phytologist* 237:1590–1605.
19. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4.
20. Chaturvedi S, Lucas LK, Buerkle CA, Fordyce JA, Forister ML, Nice CC, Gompert Z. 2020. Recent hybrids recapitulate ancient hybrid outcomes. *Nature Communications* 11.
21. Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:884–890.
22. Chen Y, Zhou R, Hu Q, Wei W, Liu J. 2021. Conservation and divergence of the Constans-like (COL) genes related to flowering and circadian rhythm in brassica napus. *Frontiers in Plant Science* 12.
23. Cheng X, Peng J, Ma J, Tang Y, Chen R, Mysore KS, Wen J. 2012. no apical meristem (mtnam) regulates floral organ identity and lateral organ separation in medicago truncatula. *New Phytologist* 195:71–84.
24. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.
25. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10.
26. De La Torre AR, Roberts DR, Aitken SN. 2014. Genome-wide admixture and ecological niche modelling reveal the maintenance of species boundaries despite long history of interspecific gene flow. *Molecular Ecology* 23:2046–2059.
27. De La Torre AR, Wang T, Jaquish B, Aitken SN. 2013. Adaptation and exogenous selection in apicea glauca×picea engelmanniihybrid zone: Implications for forest management under climate change. *New Phytologist* 201:687–699.
28. Del-Rio G, Rego MA, Whitney BM, Schunck F, Silveira LF, Faircloth BC, Brumfield RT. 2021. Displaced clines in an avian hybrid zone (thamnophilidae:rhegmatorhina) within an Amazonian Interfluve. *Evolution* 76:455–475.

29. Depardieu C, Gérardi S, Nadeau S, Parent GJ, Mackay J, Lenz P, Lamothe M, Girardin MP, Bousquet J, Isabel N. 2021. Connecting tree-ring phenotypes, genetic associations and transcriptomics to decipher the genomic architecture of drought adaptation in a widespread conifer. *Molecular Ecology* 30:3898–3917.
30. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43:491–498.
31. Emms DM, Kelly S. 2019. Orthofinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology* 20.
32. Faure S, Higgins J, Turner A, Laurie DA. 2007. Theflowering locus T-like gene family in barley (hordeum vulgare). *Genetics* 176:599–609.
33. Forester BR, Lasky JR, Wagner HH, Urban DL. 2018. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology* 27:2215–2233.
34. Gagalova KK, Warren RL, Coombe L, Wong J, Nip KM, Yuen MM, Whitehill JG, Celedon JM, Ritland C, Taylor GA, et al. 2022. Spruce giga-genomes: Structurally similar yet distinctive with differentially expanding gene families and rapidly evolving genes. *The Plant Journal* 111:1469–1485.
35. Gompert Z, Buerkle CA. 2009. A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology* 18:1207–1224.
36. Gompert Z, Buerkle CA. 2010. introgress: A software package for mapping components of isolation in hybrids. *Mol Ecol Resour*. 10:378–384.
37. Gompert Z, Buerkle CA. 2011. Bayesian estimation of Genomic Clines. *Molecular Ecology* 20:2111–2127.
38. Gompert Z, Buerkle CA. 2012. BGC: Software for bayesian estimation of Genomic Clines. *Mol Ecol Resour*. 12:1168–1176.
39. Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* 202:18–27.
40. Hamilton JA, Lexer C, Aitken SN. 2012. Differential introgression reveals candidate genes for selection across a spruce (*Picea sitchensis x P. glauca*) hybrid zone. *New Phytologist* 197:927–938.
41. Hamilton JA, Miller JM. 2015. Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conservation Biology* 30:33–41.
42. Han G, Qiao Z, Li Y, Yang Z, Wang C, Zhang Y, Liu L, Wang B. 2022. Ring zinc finger proteins in plant abiotic stress tolerance. *Frontiers in Plant Science* 13.
43. Harrison RG, Larson EL. 2014. Hybridization, introgression, and the nature of Species Boundaries. *Journal of Heredity* 105:795-809.
44. Haselhorst MS, Buerkle CA. 2013. Population genetic structure of Picea engelmannii, P. glauca and their previously unrecognized hybrids in the central Rocky Mountains. *Tree Genetics & Genomes* 9:669–681.
45. Hasson A, Blein T, Laufs P. 2010. Leaving the meristem behind: The genetic and molecular control of leaf patterning and morphogenesis. *Comptes Rendus. Biologies* 333:350–360.
46. Hedrick PW. 2013. Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology* 22:4606–4618.
47. Hijmans RJ. 2010. Raster: Geographic Data Analysis and modeling. CRAN: Contributed Packages.

48. Hornoy B, Pavy N, Gérardi S, Beaulieu J, Bousquet J. 2015. Genetic adaptation to climate in white spruce involves small to moderate allele frequency shifts in functionally diverse genes. *Genome Biology and Evolution* 7:3269–3285.
49. Horton, KW 1959. Characteristics of subalpine spruce in Alberta. Canada Department of Northern Affairs, Natural Resources, Forest Research Division, *Technical Note* 76.
50. Hu R, Qi G, Kong Y, Kong D, Gao Q, Zhou G. 2010. Comprehensive analysis of NAC domain transcription factor gene family in populus trichocarpa. *BMC Plant Biology* 10:145.
51. Jaramillo-Correa JP, Beaulieu J, Bousquet J. 2001. Contrasting evolutionary forces driving population structure at expressed sequence tag polymorphisms, allozymes and quantitative traits in white spruce. *Molecular Ecology* 10:2729–2740.
52. Jiang A, Liu X, Zhu Z, Chen M. 2024. Genome-wide identification of the AP2/ERF gene family from limonium bicolor and functional characterization of LBAP2/ERF32 under Salt Stress. *Plant Physiology and Biochemistry* 215:109035.
53. Jiang Q, Wang Z, Hu G, Yao X. 2022. Genome-wide identification and characterization of AP2/ERF gene superfamily during flower development in Actinidia eriantha. *BMC Genomics* 23.
54. Jung J, Ju Y, Seo PJ, Lee J, Park C. 2011. The SOC1-SPL module integrates photoperiod and gibberellic acid signals to control flowering time in Arabidopsis. *The Plant Journal* 69:577–588.
55. Knief U, Bossu CM, Saino N, Hansson B, Poelstra J, Vijay N, Weissensteiner M, Wolf JB. 2019. Epistatic mutations under divergent selection govern phenotypic variation in the Crow Hybrid Zone. *Nature Ecology & Evolution* 3:570–576.
56. Langdon QK, Powell DL, Kim B, Banerjee SM, Payne C, Dodge TO, Moran B, Fascinetto-Zago P, Schumer M. 2022. Predictability and parallelism in the contemporary evolution of hybrid genomes. *PLOS Genetics* 18: e1009914.
57. Laurin GV, Francini S, Penna D, Zuecco G, Chirici G, Berman E, Coops NC, Castelli G, Bresci E, Preti F, et al. 2022. SnowWarp: An open science and open data tool for daily monitoring of Snow Dynamics. *Environmental Modelling & Software* 156:105477.
58. Lechner E, Achard P, Vansiri A, Potuschak T, Genschik P. 2006. F-box proteins everywhere. *Current Opinion in Plant Biology* 9:631–638.
59. Ledig FT, Hodgskiss PD, Johnson DR. 2006. The structure of genetic diversity in engelmann spruce and a comparison with blue spruce. *Canadian Journal of Botany* 84:1806–1828.
60. Leigh D, Lischer H, Guillaume F, Grossen C, Günther T. 2021. Disentangling adaptation from drift in bottlenecked and reintroduced populations of Alpine ibex. *Mol. Ecol. Resour.* 21(7):2350–2363.
61. Levin DA, Francisco-Ortega J, Jansen RK. 1996. Hybridization and the extinction of rare plant species. *Conservation Biology* 10:10–16.
62. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760.
63. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
64. Liepe KJ, Hamann A, Smets P, Fitzpatrick CR, Aitken SN. 2016. Adaptation of lodgepole pine and interior spruce to climate: Implications for reforestation in a warming world. *Evolutionary Applications* 9:409–419.
65. Liu Y, Imai R. 2018. Function of plant dexd/h-box RNA helicases associated with ribosomal RNA biogenesis. *Frontiers in Plant Science* 9.

66. Lockwood JD, Aleksić JM, Zou J, Wang J, Liu J, Renner SS. 2013. A new phylogeny for the genus picea from plastid, mitochondrial, and nuclear sequences. *Molecular Phylogenetics and Evolution* 69:717–727.
67. Lotterhos KE, Whitlock MC. 2014. Evaluation of demographic history and neutral parameterization on the performance of fst outlier tests. *Molecular Ecology* 23:2178–2192.
68. Lotterhos KE, Whitlock MC. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology* 24:1031–1046.
69. Lotterhos KE. 2019. The effect of neutral recombination variation on genome scans for selection. G3 Genes|Genomes|Genetics 9:1851–1867.
70. Mallet J. 2005. Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* 20:229–237.
71. Martin BT, Chafin TK, Douglas MR, Douglas ME. 2021. ClineHelpR: An R package for genomic cline outlier detection and visualization. *BMC Bioinformatics* 22.
72. Matías-Hernández L, Aguilar-Jaramillo AE, Marín-González E, Suárez-López P, Pelaz S. 2014. Rav Genes: Regulation of floral induction and beyond. *Annals of Botany* 114:1459–1470.
73. McFarlane SE, Jahner JP, Lindtke D, Buerkle CA, Mandeville EG. 2024. Selection leads to remarkable variability in the outcomes of hybridisation across replicate hybrid zones. *Molecular Ecology* 33(11): e17359.
74. Nocchi G, Whiting JR, Yeaman S. 2024. Repeated global adaptation across plant species. *PNAS* 121 (52).
75. Nota F, Cambiagno DA, Ribone P, Alvarez ME. 2015. Expression and function of ATMBD4L, the single gene encoding the nuclear DNA glycosylase MBD4L in Arabidopsis. *Plant Science* 235:122–129.
76. Paul A-L, Folta KM, Ferl RJ. 2008. 14-3-3 proteins, red light, and photoperiodic flowering. *Plant Signaling & Behavior* 3:511–515.
77. Pavy N, Lamothe M, Pelgas B, Gagnon F, Birol I, Bohlmann J, Mackay J, Isabel N, Bousquet J. 2017. A high-resolution reference genetic map positioning 8.8 K genes for the conifer white spruce: Structural genomics implications and correspondence with physical distance. *The Plant Journal* 90:189–203.
78. Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
79. Raj A, Stephens M, Pritchard JK. 2014. FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* 197:573–589.
80. Rajora OP, Dancik BP. 2000. Population genetic variation, structure, and evolution in Engelmann Spruce, white spruce, and their natural hybrid complex in Alberta. *Canadian Journal of Botany* 78:768–780.
81. Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, Durphy JL, Schwarzbach AE, Donovan LA, Lexer C. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301:1211–1216.
82. Rieseberg LH, Whitton J, Gardner K. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 152:713–727.
83. Rigault P, Boyle B, Lepage P, Cooke JEK, Bousquet J, MacKay JJ. 2011. A white spruce gene catalog for Conifer Genome Analyses. *Plant Physiology* 157:14–28.
84. Roy S. 2016. Function of MYB domain transcription factors in abiotic stress and epigenetic control of stress response in plant genome. *Plant Signaling & Behavior* 11.

85. Schomburg FM, Patton DA, Meinke DW, Amasino RM. 2001. FPA, a gene involved in floral induction in Arabidopsis, encodes a protein containing RNA-recognition motifs. *The Plant Cell* 13:1427.
86. Sedghifar A, Saelao P, Begun DJ. 2016. Genomic patterns of geographic differentiation in drosophila simulans. *Genetics* 202:1229–1240.
87. Semenov G, Kenyon H, Funk E, Anderson W, McQuillan M, Spinelli J, Russell A, Martinez N, Van Huynh A, Alexander A, et al. 2025. Replicate geographic transects across a hybrid zone reveal parallelism and differences in the genetic architecture of reproductive isolation. *Evolution Letters*: qraf009.
88. Shchur V, Svedberg J, Medina P, Corbett-Detig R, Nielsen R. 2020. On the distribution of tract lengths during Adaptive Introgression. *G3 Genes|Genomes|Genetics* 10:3663–3673.
89. Simkin AJ, McAusland L, Lawson T, Raines CA. 2017. Overexpression of the Rieskefes protein increases electron transport rates and biomass yield. *Plant Physiology* 175:134–145.
90. Simon A, Fraïsse C, El Ayari T, Liautard-Haag C, Strelkov P, Welch JJ, Bierne N. 2020. How do species barriers decay? concordance and local introgression in mosaic hybrid zones of Mussels. *Journal of Evolutionary Biology* 34:208–223.
91. Suren H, Hodgins KA, Yeaman S, Nurkowski KA, Smets P, Rieseberg LH, Aitken SN, Holliday JA. 2016. Exome capture from the spruce and pine giga-genomes. *Mol Ecol Resour*. 16:1136–1146.
92. Taylor, TMC. 1959. The taxonomic relationship between Picea glauca (Moench) Voss and Picea engelmannii Parry. *Madrono* 15: 111–115.
93. VanRaden PM. 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91:4414–4423.
94. Wahl V, Ponnu J, Schlereth A, Arrivault S, Langenecker T, Franke A, Feil R, Lunn JE, Stitt M, Schmid M. 2013. Regulation of flowering by trehalose-6-phosphate signaling in arabidopsis thaliana. *Science* 339:704–707.
95. Walsh J, Billerman SM, Butcher BG, Rohwer VG, Toews DP, Vila-Coury V, Lovette IJ. 2023. A complex genomic architecture underlies reproductive isolation in a North American oriole hybrid zone. *Communications Biology* 6.
96. Wang G-L, An Y-H, Zhou C-L, Hu Z-Z, Ren X-Q, Xiong A-S. 2022. Transcriptome-wide identification of NAC (no apical meristem/arabidopsis transcription activation factor/cup-shaped cotyledon) transcription factors potentially involved in salt stress response in garlic. *PeerJ* 10.
97. Wang Y, Chen F, Chen Y, Ren K, Zhao D, Li K, Li H, Wan X, Peng M, Xiang Z, et al. 2024. Identification and analysis of drought-responsive F-box genes in upland rice and involvement of OSFBX148 in ABA response and Ros Accumulation. *BMC Plant Biology* 24.
98. Warren RL, Keeling CI, Yuen MM, Raymond A, Taylor GA, Vandervalk BP, Mohamadi H, Paulino D, Chiu R, Jackman SD, et al. 2015. Improved white spruce (picea glauca) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. *The Plant Journal* 83:189–212.
99. Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of Population Structure. *Evolution* 38:1358.
100. Westram AM, Faria R, Johannesson K, Butlin R. 2021. Using replicate hybrid zones to understand the genomic basis of adaptive divergence. *Molecular Ecology* 30:3797–3814.
101. Whiting JR, Booker TR, Rougeux C, Lind BM, Singh P, Lu M, Huang K, Whitlock MC, Aitken SN, Andrew RL, et al. 2024. The genetic architecture of

repeated local adaptation to climate in distantly related plants. *Nature Ecology & Evolution* 8:1933–1947.
102. Wimmer V, Albrecht T, Auinger H-J, Schön C-C. 2012. Synbreed: A framework for the analysis of genomic prediction data using R. *Bioinformatics* 28:2086–2087.
103. Wu TD, Reeder J, Lawrence M, Becker G, Brauer MJ. 2016. GMAP and GSNAP for genomic sequence alignment: Enhancements to speed, accuracy, and functionality. *Methods in Molecular Biology* 1418:283–334.
104. Xu J, Wang X, Guo W. 2015. The cytochrome P450 superfamily: Key players in plant development and Defense. *Journal of Integrative Agriculture* 14:1673–1686.
105. Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, Degner JC, Nurkowski KA, Smets P, Wang T, Gray LK, et al. 2016. Convergent local adaptation to climate in distantly related conifers. *Science* 353:1431–1433.
106. Yu Z, Tian C, Guan Y, He J, Wang Z, Wang L, Lin S, Guan Z, Fang W, Chen S, et al. 2022. Expression analysis of TCP transcription factor family in autopolyploids of Chrysanthemum Nankingense. *Frontiers in Plant Science* 13.
107. Zeng J, Zuo T, Liu Y, Tao H, Mo Y, Li C, Zhao L, Gao J. 2022. Phylogenetic analysis of PP2C proteins and interactive proteins analyze of BJUPP2C52 in brassica juncea. *Plant Physiology and Biochemistry* 179:25–31.
108. Zhang L, Wang P, Wang M, Xu X, Jia H, Wu T, Yuan S, Jiang B, Sun S, Han T, et al. 2024. GMTCP40 promotes soybean flowering under long-day conditions by binding to the GMAP1A promoter and upregulating its expression. *Biomolecules* 14:465.
109. Zhang X, Gou M, Guo C, Yang H, Liu C-J. 2014. Down-regulation of Kelch Domain-containing F-box protein in Arabidopsis enhances the production of (poly)phenols and tolerance to ultraviolet radiation . *Plant Physiology* 167:337–350.
110. Zhang X, Gou M, Liu C-J. 2013. arabidopsis kelch repeat F-box proteins regulate phenylpropanoid biosynthesis via controlling the turnover of phenylalanine ammonia-lyase. *The Plant Cell* 25:4994–5010.

Figures

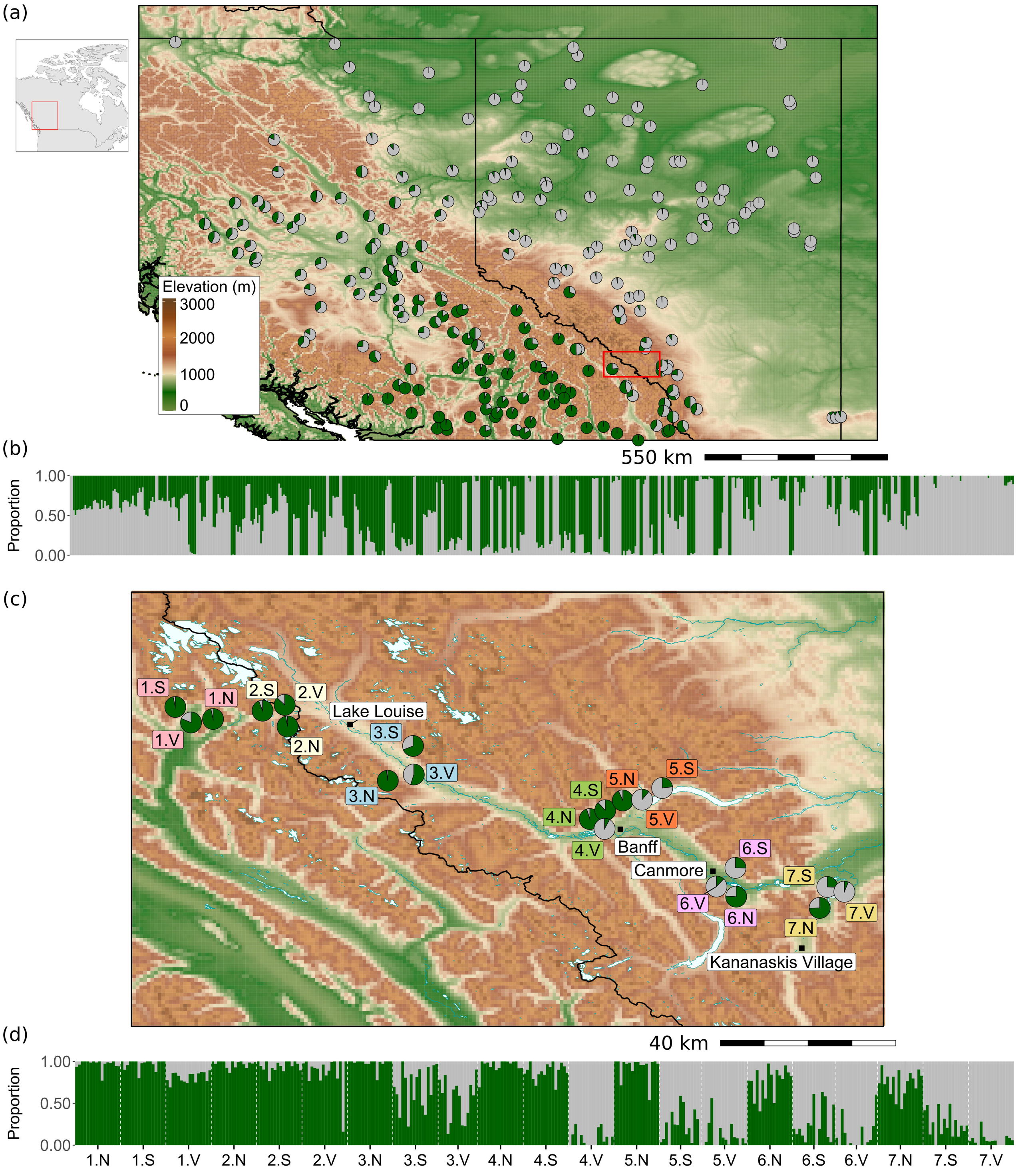
Fig. 1. Geographic locations of populations from the (a) latitudinal and (c) elevational datasets in Canada. Pie size reflects the number of individuals per population. Grey indicates *P. glauca* ancestry; green indicates *P. engelmannii* ancestry. The small overview map in (a) shows the location of the latitudinal transects in Canada. The red square in (a) marks the location of the elevational transects within the broader latitudinal hybrid zone. Panels (b) and (d) show ancestry bar plots based on fastSTRUCTURE results for the latitudinal and elevational datasets, respectively. Individual bars in (b) and (d) are ordered by population longitude.

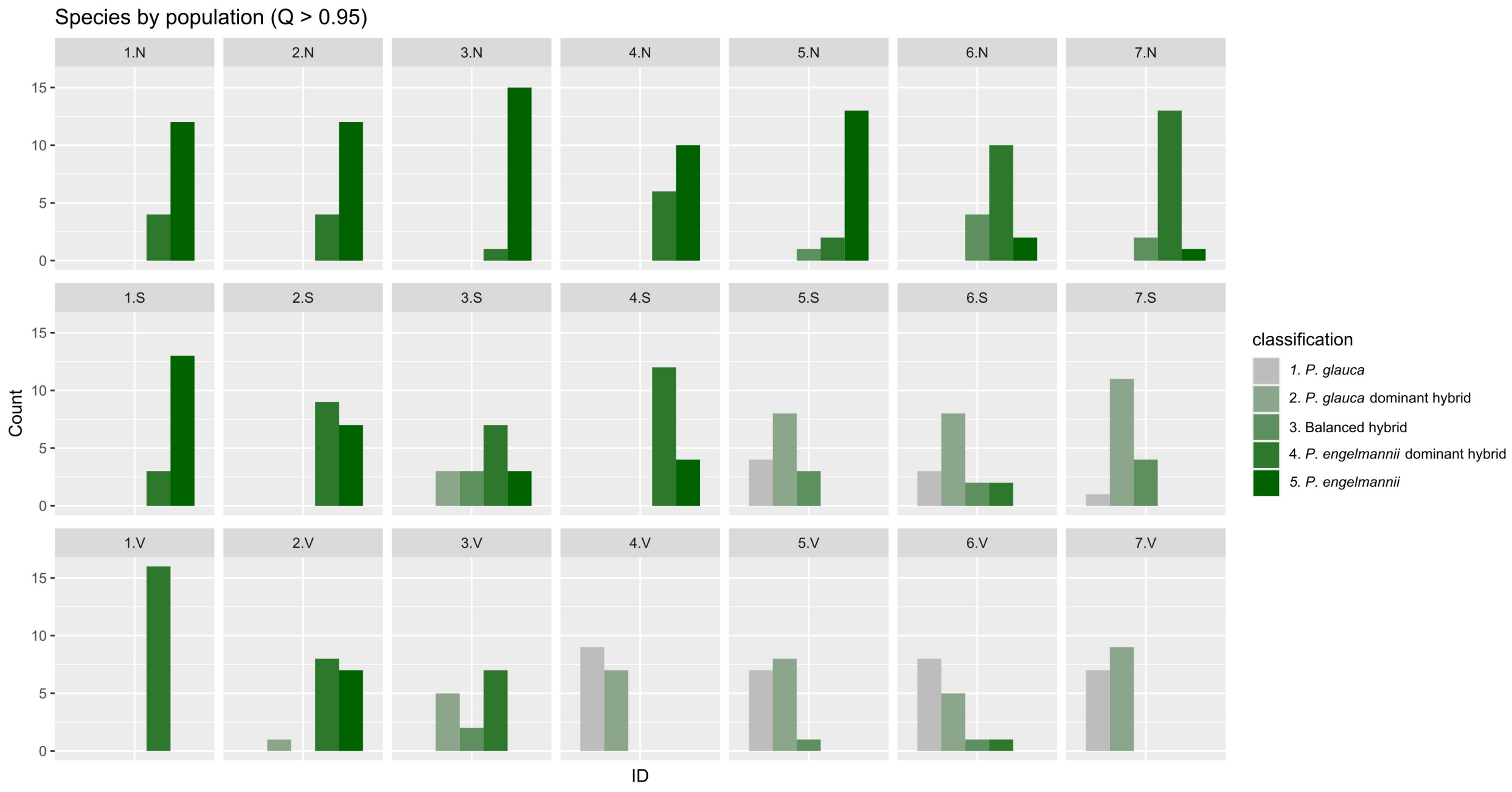
Fig. 2. Distribution of individual's ancestry category across transects and environments in the elevational datasets (N: north-facing slopes, S: south-facing slopes, V: valleys). Individuals were classified by ancestry according to the admixture coefficient computed with *fastSTRUCTURE*, using the following criteria: *P. glauca* ($q \geq 0.95$), *P. glauca* dominant hybrid ($0.60 \leq q < 0.95$), balanced hybrid ($0.40 < q < 0.60$), *P. engelmannii* dominant hybrid ($0.05 \leq q \leq 0.40$), *P. engelmannii* ($q < 0.05$).

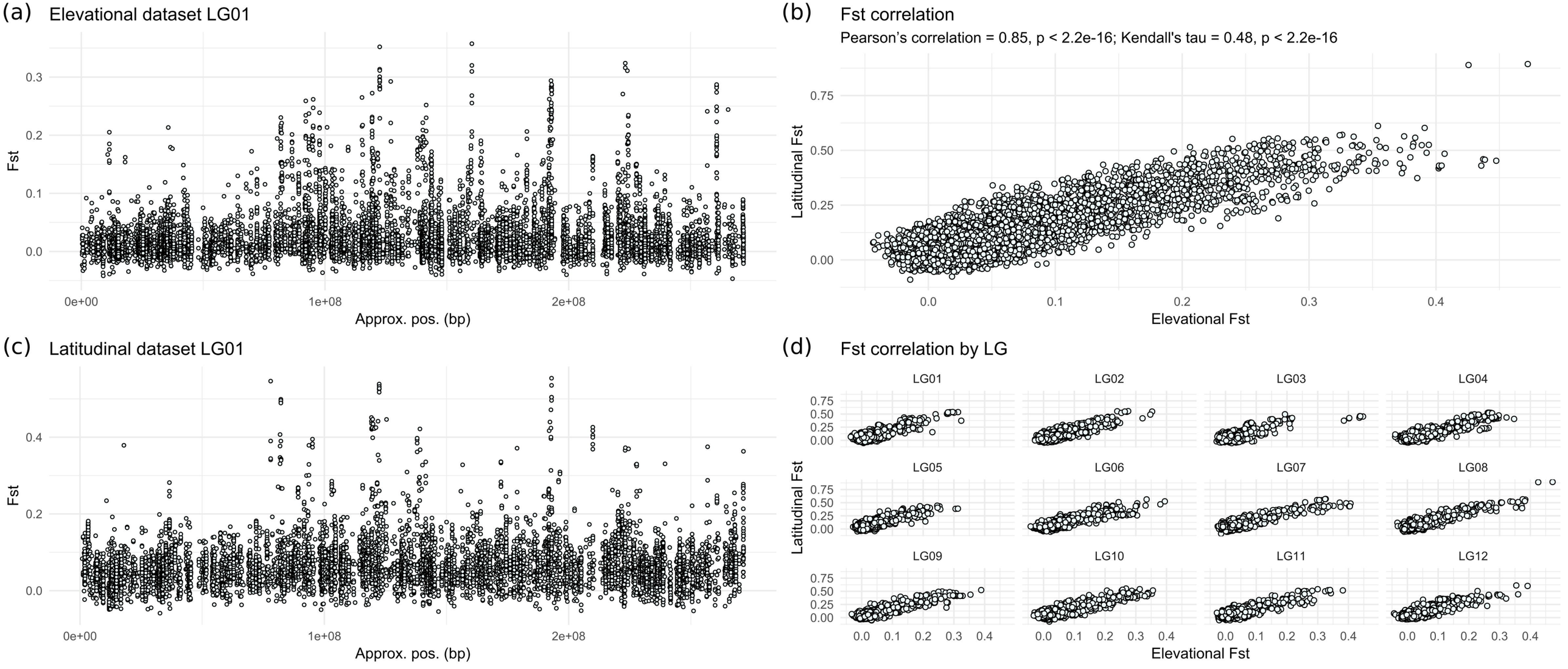
1210 **Fig. 3.** Genomic patterns of among population divergence (F_{ST}). Manhattan plots of F_{ST}
1211 values along LG01 in the (a) elevational and (c) latitudinal dataset. (b) Genome-wide
1212 correlation in F_{ST} between the shared SNPs of the elevational and latitudinal datasets. (d)
1213 Correlation in F_{ST} by linkage group (LG). Note that the significance of the p-value in (b) is
1214 overestimated as an assumption of the test is violated due to non-independence among
1215 linked SNPs.

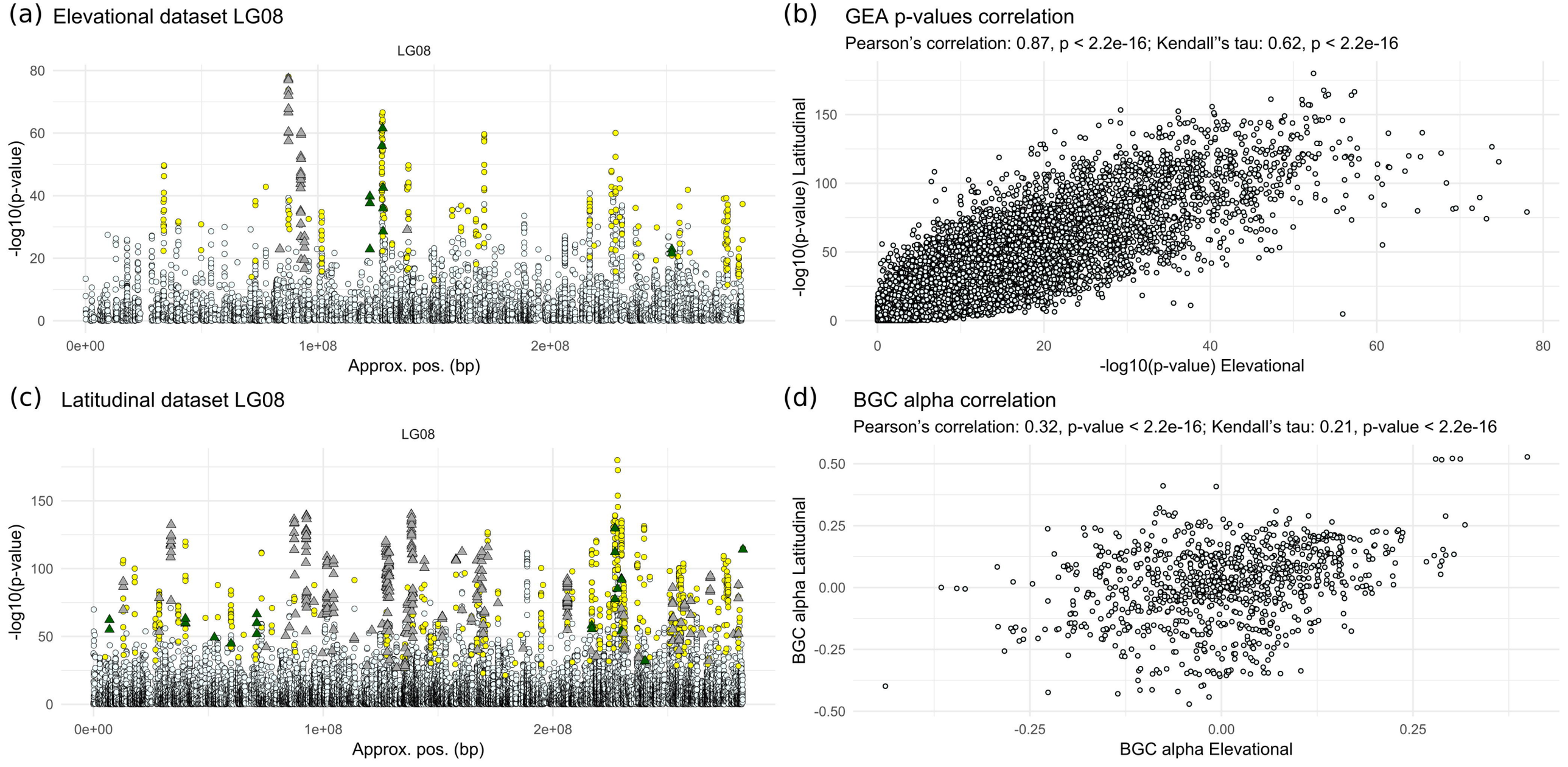
1216

1217 **Fig. 4.** Genomic patterns of Genotype-Environment Association (GEA) and genomic clines.
1218 Manhattan plots of GEA p-values along LG08 in the (a) elevational and (c) latitudinal
1219 dataset. SNPs identified as showing ancestry excess in the genomic clines (*BGC*) analyses
1220 are highlighted: dark green indicates *P. engelmannii* ancestry excess, grey indicates *P.*
1221 *glauca* ancestry excess, yellow represents non-significant SNPs tested in the *BGC* analysis,
1222 and light grey denotes SNPs not included in the *BGC* analysis. (b) Genome-wide correlation
1223 of GEA p-values between the shared SNPs of the elevational and latitudinal datasets. (d)
1224 Genome-wide correlation of BGC α values between the shared SNPs of the elevational and
1225 latitudinal datasets. Note that the significance of p-values in panels (b) and (d) is
1226 overestimated as an assumption of the test is violated due to non-independence among
1227 linked SNPs.


(a)
Elevation (m)
3000
2000
1000
0
550 km
(b)
Proportion
1.00
0.50
0.00
(c)
1.S
1.N
1.V
2.S
2.V
2.N
Lake Louise
3.S
3.V
3.N
4.S
4.N
4.V
5.N
5.S
5.V
Banff
6.S
Canmore
6.V
6.N
7.S
7.V
7.N
Kananaskis Village
40 km
(d)
Proportion
1.00
0.50
0.00
1.N
1.S
1.V
2.N
2.S
2.V
3.N
3.S
3.V
4.N
4.S
4.V
5.N
5.S
5.V
6.N
6.S
6.V
7.N
7.S
7.V


Species by population (Q > 0.95)
1.N
2.N
3.N
4.N
5.N
6.N
7.N
1.S
2.S
3.S
4.S
5.S
6.S
7.S
1.V
2.V
3.V
4.V
5.V
6.V
7.V
Count
ID
0
5
10
15
classification
1. P. glauca
2. P. glauca dominant hybrid
3. Balanced hybrid
4. P. engelmannii dominant hybrid
5. P. engelmannii


(a)
Elevational dataset LG01
Fst
0.3
0.2
0.1
0.0
0e+00
1e+08
2e+08
Approx. pos. (bp)
(b)
Fst correlation
Pearson's correlation = 0.85, p < 2.2e-16; Kendall's tau = 0.48, p < 2.2e-16
Latitudinal Fst
0.75
0.50
0.25
0.00
0.0
0.1
0.2
0.3
0.4
Elevational Fst
(c)
Latitudinal dataset LG01
Fst
0.4
0.2
0.0
0e+00
1e+08
2e+08
Approx. pos. (bp)
(d)
Fst correlation by LG
LG01
LG02
LG03
LG04
LG05
LG06
LG07
LG08
LG09
LG10
LG11
LG12
Latitudinal Fst
0.75
0.50
0.25
0.00
0.0
0.1
0.2
0.3
0.4
Elevational Fst


(a) Elevational dataset LG08
LG08
-log10(p-value)
Approx. pos. (bp)
0e+00
1e+08
2e+08
(b) GEA p-values correlation
Pearson's correlation: 0.87, p < 2.2e-16; Kendall''s tau: 0.62, p < 2.2e-16
-log10(p-value) Latitudinal
-log10(p-value) Elevational
(c) Latitudinal dataset LG08
LG08
-log10(p-value)
Approx. pos. (bp)
0e+00
1e+08
2e+08
(d) BGC alpha correlation
Pearson's correlation: 0.32, p-value < 2.2e-16; Kendall's tau: 0.21, p-value < 2.2e-16
BGC alpha Latitudinal
BGC alpha Elevational