Replicated hybrid zones reveal genomic patterns of local adaptation and introgression in spruce

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

Hybridization between species can occur along repeated zones of contact, providing a natural laboratory for studying the interplay between migration and selection, and 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 hybrid zones derived from the same parental species are not evolutionarily independent. Conversely, processes

that operate within the middle of each hybrid zone, such as selection driving directional introgression, may be more evolutionarily independent, and therefore provide stronger evidence of selection. Here we compared hybridization and local adaptation patterns between two replicated regions within the western Canada interior spruce hybrid zone: a broad latitudinal transect with gradual environmental variation and a narrow elevational transect with substantial topographical and environmental variation. We discovered a complex pattern of introgression, with strong differences in ancestry maintained even across small spatial scales at several locations along the elevational transect. Despite differences in their spatial scales, the elevational and latitudinal transects revealed strikingly similar genome-wide patterns of differentiation and adaptation, and consistent patterns of directional introgression. We explore the extent to which the evolutionary non-independence of these hybrid zones allows inferences about the role of natural selection and drift in shaping these patterns. Consistent with theory, we found longer genomic tracts in the elevational transect, likely because the steeper environmental gradients over shorter distances limit the rate of mixing by migration and recombination relative to drift and selection.

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 has also been documented in many animal species (Hedrick 2013; Aguillon et al. 2022). Hybridization and subsequent introgression can have substantial impacts on adaptation, evolvability and speciation, with effects that may be beneficial or deleterious 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 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 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 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 can lead to similar genotype-environment associations (GEAs), whereas drift will result in differing patterns of association. As long as the repeated gradients are evolutionarily independent, repeated GEA signals can provide evidence of a gene having evolved in response to selection (Whiting et al. 2024). However, if two previously isolated species come into repeated contact across multiple hybrid zones, these zones are fundamentally non-independent, as they are repeatedly derived from the same parental species pools. If the parental species diverged in allopatry

followed by secondary contact in repeated hybrid zones, loci that had differentiated by genetic drift could exhibit repeated associations with the overall environmental gradient that distinguishes the parental species. Thus, repeatability in GEA signals across multiple repeated 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 direction across multiple hybrid zones, this would be unlikely to occur due to the above-described non-independence, and would implicate natural selection (Rieseberg et al. 1999; Buerkle and Lexer 2008; Gompert and Buerkle 2009; Bierne et al. 2013; Simon et al. 2020; Westram et al. 2021; Langdon et al. 2022). Our aim is to explore the extent of similarity in genomic patterns of introgression in two hybrid zones spanning regions with very different geography: one a broad plateau, and another a narrow and rugged mountainous region. In both regions, introgression occurs between white spruce (*Picea glauca*) and Engelmann spruce (*Picea engelmannii*), but in the former region, the environment distinguishing the two parental species transitions slowly across space, whereas in the latter region, the environmental transition is much narrower and includes considerable fine-scale variation associated with elevation and slope aspect.

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). De La Torre et al. (2014) 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 and exhibit higher fitness in habitats that are ecologically 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 in interior spruce is well supported by common garden experiments as well as 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 study hybridization and local adaptation we compare genomic patterns across two datasets: a previously studied broad latitudinal gradient spanning > 1000km throughout

133 Alberta and British Columbia (Fig. 1a) (Yeaman et al. 2016), and a new dataset collected
134 across a narrower elevational zone spanning ~100km in the Canadian Rocky Mountains
135 (Fig. 1c). Based on casual observations of the elevational zone, forests on high north-facing
136 slopes tend to have moister soil and hold snow longer in the springtime than those on high
137 south-facing slopes and valley bottoms, so we expected this might generate some fine-
138 grained natural selection within each location. To test whether this relatively fine-scale
139 environmental variation could drive adaptation, at seven locations along the valleys along
140 the elevational zone, we sampled from three distinct environments: valley bottoms (< 1600
141 m), and high elevation north- and south-facing slopes (> 2000 m). To compare the latitudinal
142 and elevational datasets, we analyse F_{ST} to characterize patterns of genomic divergence,
143 conduct genotype-environment association (GEA) analyses to identify regions potentially
144 driving local adaptation, and perform genomic cline analyses to detect regions showing
145 abnormal patterns of introgression. We also explore patterns of chromosomal clustering of
146 these statistics in each of the datasets to test whether they are consistent with expectations
147 based on theory of hybrid zones and migration-selection balance. We aim to answer three
148 questions: 1) does local adaptation maintain species divergence across fine-scale
149 environmental variation within the hybrid zone; 2) how similar are genomic patterns of
150 divergence, environmental association, and directional introgression across these very
151 different geographies of hybridization?; and 3) to what extent can we parse evolutionary non-
152 independence of these hybrid zones to identify regions of the genome driving adaptive
153 divergence?

154

155 # Results

156 **Super-scaffold assembly**

157 While the reference genome assembly of *Picea glauca* is highly fragmented (Warren et al.
158 2015), we used a previously constructed linkage map based on cDNA sequences to position

159 contigs onto chromosome scale scaffolds, to facilitate study of genome-wide patterns. To
160 maximize confidence, we identified the best scaffold match for each cDNA based on
161 sequence coverage and identity and discarded unmapped cDNA as well as those with low-
162 quality mapping or unresolved multi-mapping (see *Methods* for details). This process
163 retained 21,252 cDNAs, of which 11,914 were included in the available *P. glauca* linkage
164 map, including 14,727 genes spread across 12 linkage groups (LGs) (Pavy et al. 2017;
165 Gagalova et al. 2022). These 11,914 cDNAs mapped to 10,475 unique scaffolds of the *P.*
166 *glauca* reference genome, which we ordered along the 12 linkage groups according to the
167 cDNAs centimorgan (cM) positions in the linkage map. The resulting super-scaffold
168 assembly spanned 3,426,128,031 bp (~3.42 Gbp), representing approximately 13% of the
169 WS77111v2 *P. glauca* reference genome assembly (~26.5 Gbp). The final size of our
170 assembly is comparable to a recently published super-scaffold assembly for *P. glauca*, which
171 achieved a size of ~3.86 Gbp by using a less conservative cDNA mapping approach
172 (Gagalova et al. 2022).

173

174 **Sampling and SNP Discovery**

175 We sampled 331 spruce individuals, including white spruce (*Picea glauca*), Engelmann
176 spruce (*Picea engelmannii*), and their hybrids, from 21 populations within a narrow section of
177 the western Canada interior spruce hybrid zone which extends from central British Columbia
178 (B.C.) into western Alberta (Fig. 1c). Samples were collected from seven locations along the
179 Trans-Canada Highway going through Canmore, Banff, and Yoho. Each location included
180 three populations, with 14 to 16 individuals sampled per population (supplementary fig. S1,
181 Supplementary Material online). For each location, we sampled a north-facing high-altitude
182 population (N), a south-facing high-altitude population (S), and a valley population (V) (Fig.
183 1c; supplementary fig. S1, Supplementary Material online). Using a target sequencing
184 approach with a 10 Mb target, we identified a set of 119,664 SNPs (MAF > 0.05) following
185 post-processing. This set included only SNPs located on the scaffolds of the *P. glauca*

186 reference genome that we were able to anchor to the *P. glauca* linkage map (Gagalova et al.
187 2022) in our super-scaffold assembly construction. We termed this dataset the "elevational
188 dataset".

189 Additionally, we gathered available exome capture data for 575 spruce individuals from 249
190 populations, derived from previous studies (Suren et al. 2016; Yeaman et al. 2016). This
191 dataset spans both species' geographic ranges in B.C. and Alberta and encompasses the
192 entire western Canada interior spruce hybrid zone, with the exclusion of areas within
193 national parks (Fig. 1a; supplementary fig. S1, Supplementary Material online). After
194 applying the same SNP-calling pipeline and post-processing as for the elevational dataset to
195 ensure consistency, we identified a set of 265,655 SNPs (MAF > 0.05) located on the
196 scaffolds of our super-scaffold assembly. We termed this dataset the "latitudinal dataset".

197

198 **Population structure, ancestry, and fine-scale adaptation**

199 We analysed genome-wide variation in both elevational and latitudinal transects using
200 principal component analysis (PCA) and a variational Bayesian framework implemented in
201 *fastSTRUCTURE* (Raj et al. 2014). In both datasets, *fastSTRUCTURE* identified K=2 as the
202 number of clusters that maximizes the log marginal likelihood (K*ε) of the data (Figs. 1b and
203 d; supplementary fig. S2, Supplementary Material online). Concordantly, PCA's first principal
204 component (PC1) revealed two main clusters of individuals present in both datasets, while
205 capturing a large portion of the variation in the data (70% in the latitudinal dataset and 63%
206 in the elevational dataset) as indicated by the sharp drop in eigenvalue scores after PC1
207 (supplementary figs. S3 and S4, Supplementary Material online).

208 To assign the two clusters identified with *fastSTRUCTURE* and PCA to their respective
209 species, we added publicly available genomic data of a pure individual of *P. glauca* and one
210 of *P. engelmannii* to each dataset, retrieved from the European Nucleotide Archive (ENA)
211 (see *Methods* for details). By repeating the *fastSTRUCTURE* analysis at K=2, we could

212 clearly identify which species ancestry each cluster represented based on the pure
213 individuals admixture coefficients (supplementary fig. S5, Supplementary Material online).
214 Overall, PCA and *fastSTRUCTURE* effectively distinguished the different genomic
215 background of the two spruce species and produced highly concordant results: the
216 distribution of spruce individuals along PC1 reflected the arrangement of the individuals
217 based on the ancestry coefficients estimated with *fastSTRUCTURE* at K=2, with a nearly
218 perfect correlation (0.99) in both datasets (supplementary fig. S6, Supplementary Material
219 online). In PCA, the remaining principal components (PC2, PC3) explained considerably less
220 variation than PC1 in both the elevational and latitudinal transects and separated a few
221 second- and third-degree relatives found within some populations from the rest of the
222 individuals (supplementary figs. S3 and S4, Supplementary Material online). Overall,
223 genomic relatedness levels were low according to our estimation, with only two potential
224 half-sibling relationships (kin ~0.25) in the elevational dataset and ten in the latitudinal
225 dataset (supplementary fig. S7, Supplementary Material online). Both datasets showed few
226 third-degree relationships (supplementary fig. S7, Supplementary Material online).

227 The elevational transect exhibited two dominant patterns of covariation between ancestry
228 and environment: 1) ancestry varied from almost pure Engelmann to almost pure white
229 spruce along the broad NW to SE transect across the sampling region (spanning ~100 km of
230 distance along the valley); and 2) within several locations through the middle of the
231 elevational hybrid zone (particularly locations 4, 5, and 6; Fig. 1c), the proportion of
232 Engelmann ancestry was highest in high elevation N-facing sites (N), lowest in valleys (V),
233 and intermediate on high elevation S-facing sites (Figs. 1c and 1d; Fig. 2; supplementary fig.
234 S8, Supplementary Material online). Within each location, the N, S, and V sites were never
235 separated by more than 10km, and each location was typically separated by ~5-30 km.
236 Thus, instead of seeing a gradual transition in hybrid ancestry across the elevational transect
237 (with similar ancestry among N, S, and V within locations), we observed substantial fine-
238 grained local adaptation within locations (Fig. 2; supplementary fig. S8, Supplementary

239 Material online). Given that similar changes in ancestry are seen over ~500km in the
240 latitudinal transect, this represents a comparatively sharp transition in ancestry across micro-
241 environmental variations, maintained over several sampling locations.

242 For the latitudinal dataset, ancestry distribution aligned well with the geographical ranges of
243 the two species in western Canada (Figs. 1a and 1b) (Rajora and Dancik 2000; De La Torre
244 et al. 2013, 2014; Gagalova et al. 2022). Potentially pure *P. glauca* and *P. glauca*-dominant
245 hybrids were found predominant in the northern and eastern part of the sampling range
246 going into Alberta, while *P. engelmannii* ancestry appeared more prevalent in south-western
247 British Columbia. More balanced species density with varying admixture levels were
248 observed in central and western B.C. and along the southern B.C.-Alberta border, matching
249 with the extent of what is commonly referred to as the western Canada interior spruce hybrid
250 zone (Fig. 1a).

251

252 **Population differentiation**

253 We measured the extent of genetic differentiation between populations in each dataset using
254 the F_{ST} statistic on a per-SNP basis, implemented in *VCFtools* (Danecek et al. 2011). In the
255 elevational dataset, mean F_{ST} between the 21 populations sampled was 0.03 (median =
256 0.01, sd = 0.05), based on 119,664 SNPs (supplementary fig. S9, Supplementary Material
257 online). In the latitudinal dataset, mean F_{ST} between 249 populations was 0.07
258 (median=0.06, sd=0.07), based on 265,655 SNPs (supplementary fig. S9, Supplementary
259 Material online). For reference, mean F_{ST} between *P. glauca* and *P. engelmannii*, estimated
260 by selecting pure species individuals based on the *fastSTRUCTURE* admixture coefficient Q
261 ($Q \geq 0.95$), was 0.10 (median = 0.04, sd = 0.15) in the elevational dataset and 0.14 (median
262 = 0.07, sd = 0.17) in the latitudinal dataset (supplementary fig. S9, Supplementary Material
263 online). Most SNPs were polymorphic within both pure species ($Q \geq 0.95$) populations,
264 indicating substantial shared variation despite genetic differentiation. Specifically, in the

265 elevational dataset (119,664 SNPs), 96.7 % of SNPs (115,768 SNPs) were polymorphic in *P.*
266 *glauca* and 99.7 % (119,415 SNPs) in *P. engelmannii*; in the latitudinal dataset (265,655
267 SNPs), 98.6% of SNPs (262,023 SNPs) were polymorphic in *P. glauca* and 96% (255,500
268 SNPs) in *P. engelmannii*.

269 Comparison of the SNPs analysed in each dataset revealed that 31,667 SNPs were shared.
270 This shared set was relatively limited, primarily due to differences in the sequence capture
271 designs used for the two datasets. Moreover, small population sizes and missing data in the
272 latitudinal dataset prevented F_{ST} computation for several loci, restricting the cross-dataset
273 F_{ST} comparisons to 11,990 SNPs (out of the 31,667 shared SNPs). For this subset of 11,990
274 SNPs, the correlation of F_{ST} values was high (Pearson's $r = 0.85$, $p < 2.2 \times 10^{-16}$; Kendall's
275 tau = 0.48, $p < 2.2 \times 10^{-16}$), with little variation observed among linkage groups (Fig. 3).
276 Overall, the F_{ST} landscape between the elevational and latitudinal transects appears highly
277 repeated, with a visually concordant distribution of peaks and valleys in the resulting
278 Manhattan plots (Fig. 3a and c; supplementary fig. S10, Supplementary Material online).
279 Patterns of nucleotide diversity among parental populations (d_{xy}) were broadly concordant
280 with F_{ST}, and we observed relatively low correlation between nucleotide diversity within each
281 parental species (π) and F_{ST}, with only a slight tendency for the most extreme F_{ST} outliers to
282 have reduced diversity (supplementary figs. S11 and S12, Supplementary Material online).
283 This suggests that background selection reducing π is not the main driver of high F_{ST} (as per
284 Cruickshank and Hahn 2014).

285

286 **Genome-environment association analysis**

287 To identify the particular combination of environments that covaries most strongly with
288 variation in ancestry, we performed redundancy analysis (RDA) by regressing the individual
289 trees' hybrid indexes, computed with *fastSTRUCTURE*, against a set of environmental
290 variables considered important in describing the niches of the two parental species. This

291 approach involves a specific type of multiple regression aimed at identifying the linear
292 combination of environmental variables that accounts for the largest portion of variation in
293 the hybrid index, and therefore best explains the introgression patterns that we observed.
294 RDA1 (the first axis of RDA variation i.e. the linear combination of predictors explaining most
295 hybrid index variance) was then used as a predictor in genome-environment association
296 analyses (GEAs). For the elevational dataset, we included six weakly correlated (Pearson's
297 correlation coefficient < 0.7) environmental variables: aspect (the compass direction that a
298 terrain faces), slope (the steepness of the terrain), air temperature at 2 meters, diffuse
299 horizontal irradiation (DIF), predicted photovoltaic electricity output (PVOUT), and snow melt
300 day (supplementary fig. S13, Supplementary Material online). For the latitudinal dataset, we
301 used the same variables except snow melt day. The latter, being high-resolution data, was
302 not feasible to generate for the vast geographic area covered by the latitudinal dataset.

303 In the elevational dataset, the first RDA axis (RDA1) explained 73% of the variation in the
304 hybrid index, with snow melt day and PVOUT being the most important variables, followed
305 by air temperature at 2 meters, slope, aspect, and DIF (supplementary fig. S14,
306 Supplementary Material online). In contrast, in the latitudinal dataset RDA1 explained 56% of
307 the total hybrid index variation. The most influential variables were slope and PVOUT,
308 followed by air temperature at 2 meters and DIF (supplementary fig. S14, Supplementary
309 Material online). In this dataset, aspect contributed negligibly to the introgression patterns
310 observed across latitude, in strong contrast with its contribution in the elevational dataset
311 (supplementary fig. S14, Supplementary Material online).

312 We used the resulting RDA1 axes as predictors in the GEA, conducting separate analyses
313 for the elevational and latitudinal datasets. This analysis should detect candidates driving
314 local adaptation in the parental species, but it also may pick up any regions of the genome
315 that happened to evolve strong differentiation due to drift or intrinsic reproductive isolating
316 mechanisms. The elevational dataset analysis included 119,664 SNPs, while the latitudinal
317 included 265,655 SNPs (MAF > 0.05). For each SNP, we performed a likelihood-ratio test

comparing two models: a null model including only the intercept, where the SNP frequency was predicted using only the mean frequency; and a predictor model which used RDA1 as the environmental predictor to infer SNP frequency. The likelihood-ratio test was used to compare the goodness of fit between the predictor model and the intercept-only model. A significantly better fit by the predictor model results in a low p-value, reflecting the predictor's importance. Conversely, if the null model is supported by the observed data, the likelihoods of the two models would not differ beyond the sampling error, resulting in higher p-values. Similar to the F_{ST} landscape, the GEA landscape between the elevational and latitudinal transects appeared highly repeated, with a visually strikingly concordant distribution of peaks and valleys in the Manhattan plots derived from the two GEAs (Fig. 4a and c; supplementary fig. S15, 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: 0.84, $p < 2.2 \times 10^{-16}$; Kendall's tau = 0.58, $p < 2.2 \times 10^{-16}$), with little variation between linkage groups (Fig.4b; supplementary fig. S16, Supplementary Material online). We also tested whether the inclusion of the snow melt day variable in the elevation environmental set but not in the latitudinal set contributed to our results, but found minimal impact (supplementary Text S1 and supplementary fig. S17, Supplementary Material online).

Concordant with the F_{ST} analysis, we assessed correlations between GEA p-values and nucleotide diversity within and between parental species in each dataset. We observed strong correlations with d_{xy} and low correlations with nucleotide diversity within parental species, with only a slight tendency for the most extreme GEA outliers to have reduced diversity (supplementary figs. S11 and S12, Supplementary Material online).

The above results do not include any correction for population structure, so we explored the impact of adding a structure correction using PC1 from our PCA analysis (see *Methods* for details). We found that the resulting p-values appear entirely uncorrelated between the elevational and latitudinal analyses with almost no shared outliers, in strong contrast with

345 those of our GEA results without population structure correction (supplementary fig. S18,
346 Supplementary Material online). These observations are consistent with the expected
347 behaviour of the methods: GEA with structure correction should appropriately control the
348 proportion of false positives at the cost of low power to detect true positives, whereas GEA
349 without structure correction should be more sensitive to detecting true positives at the cost of
350 a large number of false positives (Lotterhos 2019; Booker et al. 2023). Both issues are
351 expected to be particularly pronounced when the axis of population structure is confounded
352 with environmental variation driving local adaptation, as is the case in this study. Given that
353 previous studies have shown considerable signatures of local adaptation, it seems likely that
354 GEA without correction (Fig. 4a and c; supplementary fig. S15, Supplementary Material
355 online) is capturing some true signals, along with an unknown proportion of false positives.

356

357 **Chromosomal clustering of F_{ST} and GEA signatures**

358 To examine chromosomal clustering of signatures of divergence and environmental
359 association, we calculated Moran's I index of spatial autocorrelation on chromosomal
360 distance. For clustering of F_{ST} we used 11,990 shared SNPs for each scaffold and linkage
361 group (see *Methods* for details). Although not statistically significant (scaffolds paired t-test
362 p-value = 0.49, n = 27; LGs paired t-test p-value = 0.22; n = 12), this analysis showed that
363 F_{ST} chromosomal autocorrelation tended to be slightly higher in the elevational than in the
364 latitudinal transects when assessed by LG (Elevational scaffolds Moran's I - Latitudinal
365 scaffolds Moran's I mean and median: 0.007 and 0.005; Elevational LGs Moran's I -
366 Latitudinal LGs Moran's I mean and median: 0.006 and 0.002) (supplementary fig. S19,
367 Supplementary Material online). However, the sample size of only 12 LGs doesn't allow for
368 much power, and no strong trend was detected in the scaffolds (supplementary fig. S19,
369 Supplementary Material online). We repeated this analysis on the GEA p-values from the
370 31,667 shared SNPs for each scaffold and linkage group. Levels of GEA p-values
371 chromosomal autocorrelation within scaffolds and LGs were significantly ($p < 0.05$) higher

372 (scaffolds paired t-test p-value = 7.48×10^{-5}, n = 77; LGs paired t-test p-value = 0.01, n =
373 12) in the elevational than latitudinal dataset (Elevational scaffolds Moran's I - Latitudinal
374 scaffolds Moran's I mean and median: 0.037 and 0.028; Elevational LGs Moran's I -
375 Latitudinal LGs Moran's I mean and median: 0.01 and 0.01) (supplementary fig. S20,
376 Supplementary Material online). Additionally, we tested whether the differences in GEA p-
377 values spatial autocorrelation could be driven by the distinct sampling schemes used in the
378 elevational and latitudinal datasets (15-16 individuals per population vs. 2 individuals per
379 population), and found no evidence that variation in populations sampling intensity
380 contributed to these differences (supplementary Text S2 and supplementary fig. S21,
381 Supplementary Material online).

382

383 **Genomic clines analysis of aberrant patterns of introgression**

384 Genomic clines analysis is used to infer the movement of ancestry blocks into different
385 genomic backgrounds and to assess patterns of introgression in hybrid zones, by modelling
386 the ancestry probabilities of SNPs as a function of the hybrid index. Here, we aim to identify
387 regions of the genome exhibiting extreme patterns of introgression, which might occur if
388 some loci are more strongly introgressed into one species or the other due to their adaptive
389 effect (hereafter, directional), or if some loci experience unusually steep or gradual
390 transitions across the hybrid index compared to the genome-wide average (hereafter,
391 steepness). Directional introgression (α) occurs when an allele moves from one parental
392 species into the other more often than expected based on genome-wide patterns, which
393 typically happens when a locus experiences uniform directional selection across both
394 species and hybrids (e.g. a classic selective sweep). On the other hand, steepness (β)
395 reflects how rapidly ancestry changes across the hybrid index and can arise from selection
396 favouring heterozygotes (i.e. via complementation of deleterious recessive alleles) or
397 disfavouring heterozygotes (i.e. due to reproductive incompatibilities), resulting in loci that
398 are introgressed more slowly or more rapidly than the genome-wide average (Gompert and

399 Buerkle 2009, 2010, 2011, 2012). We applied genomic clines analyses to assess the
400 patterns of introgression between *P. glauca* and *P. engelmannii* in the western Canada
401 interior spruce hybrid zone. We fit genomic clines using two approaches: a Bayesian
402 genomic clines model implemented in *BGC* (Gompert and Buerkle 2011, 2012) and a logistic
403 regression-based model implemented in *introgress* (Gompert and Buerkle 2010). We used
404 *introgress* to have a source of comparison and verify the consistency of *BGC*, but only
405 results from *BGC* were analysed further.

406 In order to define parental pure species populations for the genomic clines model, we
407 classified the sampled individuals as either *P. glauca*, *P. engelmannii* or hybrids according to
408 the admixture coefficient computed with *fastSTRUCTURE* (see *Methods* for details). We
409 performed two separate genomic clines analyses, one for the elevational and one for the
410 latitudinal dataset respectively and included in the models only SNPs with an allele
411 frequency differential between parental “pure” species populations greater than 0.6. Limiting
412 the number of SNPs tested in *BGC* was necessary, as using the full SNP set would lead to
413 infeasible run times due to the computational demands of its Bayesian framework.
414 Therefore, we opted to include only ancestry-informative SNPs (those with substantial
415 parental frequency differentials - 0.6), rather than a random subset that might contain many
416 uninformative loci, as suggested in other studies (Knief et al. 2019; Chaturverdi et al. 2020;
417 Del-Rio et al. 2021; Walsh et al. 2023) and following recommendations in Martin et al.
418 (2021). In our *BGC* analysis, positive α values indicate excess *P. glauca* ancestry compared
419 to the genome-wide average, while negative α values indicate excess *P. engelmannii*
420 ancestry (directional introgression patterns). On the other hand, positive β values indicate a
421 steeper cline and selection against introgression resulting in a faster rate of allelic-frequency
422 transition along the hybrid index, whereas negative β values indicate the reverse.

423 In the elevational dataset, pre-processing resulted in a set of 3,256 SNPs for genomic clines
424 analysis, which in turn led to the identification of 300 SNPs exhibiting excess ancestry (148
425 positive α and 152 negative α SNPs). These loci represent SNPs with introgression patterns

426 different from the genome-wide average, with a 95% equal-tail probability interval (ETPI) for
427 the posterior probability distribution of α that did not overlap with zero. No SNP showing
428 unusual rates of allelic-frequency transition along the hybrid index was detected in the
429 elevational dataset genomic clines analysis (β excess) (supplementary figs. S22 and S23,
430 Supplementary Material online).

431 In the latitudinal dataset, pre-processing resulted in a set of 10,154 SNPs for genomic clines
432 analysis, which in turn led to the identification of 3,845 SNPs with a 95% ETPI for the
433 posterior probability of α that did not overlap 0 (1,906 positive α and 1,939 negative α
434 SNPs), indicating excess ancestry (supplementary figs. S23 and S24, Supplementary
435 Material online). Additionally, the latitudinal dataset analysis pinpointed 120 SNPs exhibiting
436 unusual rates of allele frequency transitions in the hybrid population (52 negative β and 68
437 positive β SNPs) (supplementary fig. S24, Supplementary Material online). Similarly to SNPs
438 exhibiting excess ancestry (α excess SNPs), these SNPs reported a 95% ETPI for the
439 posterior probability distribution of β that did not overlap with zero (β excess). Among the
440 SNPs that reported α or β excess in the latitudinal dataset analysis, 76 were in common,
441 indicating loci exhibiting both excess ancestry as well as unusual rates of allelic-frequency
442 transitions along the hybrid index gradient (supplementary fig. S24, Supplementary Material
443 online).
444 Furthermore, we performed two additional *BGC* analyses on subsets of the latitudinal
445 dataset, referred to as Subset 1 and Subset 2, by down sampling the data to a similar
446 number of individuals and parental population composition as in the elevational dataset
447 (supplementary fig. S25, Supplementary Material online). This was done to determine
448 whether the high number of α-excess SNPs observed in the latitudinal dataset, relative to
449 the total number of SNPs tested, was a statistical artifact of their differing sample sizes, or
450 indicated a stronger biological signal in the latitudinal transects compared to the elevational
451 transects. Both downsampled analyses revealed a pattern closely resembling that of the full
452 dataset analysis, identifying 3,258 and 3,268 α-excess SNPs, respectively (supplementary

453 fig. S26, Supplementary Material online), suggesting a stronger biological signal in the
454 latitudinal dataset. Furthermore, the Pearson's correlation between α values from the full
455 latitudinal dataset analysis and those from the subset analyses was very high, at 0.91 and
456 0.90. For comparison, α correlation between the two subset analyses was even higher, at
457 0.97.

458 Comparing across datasets, the elevational and latitudinal dataset clines analyses shared a
459 total of 982 SNPs. These showed significant correlation of α values between the elevational
460 and latitudinal dataset analyses (Pearson's correlation: 0.32, p-value < 2.2×10^{-16}; Kendall's
461 tau: 0.21, p-value < 2.2×10^{-16}) (Fig. 4d). In more detail, within the 982 shared SNPs, 438
462 exhibited excess ancestry in the latitudinal dataset analysis and 83 in the elevational -
463 between these, 53 reported excess ancestry across both analyses, with 49 (92%) exhibiting
464 excess ancestry in the same direction of introgression. Overall, *BGC* α values were
465 generally in agreement with the p-value obtained with *introgress* (between *introgress* p-
466 values and *BGC* absolute α values) for both datasets (supplementary fig. S27,
467 Supplementary Material online).

468 By visually inspecting Manhattan plots derived from our genomic clines analyses, it is
469 apparent that some of the strongest GEA peaks comprising our top candidate for
470 environmental association show asymmetric introgression from one parental species to the
471 other (excess ancestry) (Fig. 4a and c; supplementary fig. S28, Supplementary Material
472 online). This pattern is further confirmed by the relationship between GEA p-values and *BGC*
473 α-values, with numerous α-outlier SNPs exhibiting low GEA p-values (supplementary fig.
474 S23, Supplementary Material online). This could indicate local adaptation associated with
475 the genomic region's ancestry, even though we cannot totally exclude stable genetic
476 differentiation (i.e. genetic drift). SNPs with these asymmetric patterns tend to be organized
477 into large blocks that all tend to be introgressed in the same direction (Fig. 4a and c;
478 supplementary fig. S28, Supplementary Material online).

479

480 **Candidate Genes for Local Adaptation**

481 Given that common garden experiments have shown these species are strongly locally
482 adapted (Jaramillo-Correa et al. 2001; De La Torre et al. 2013; Liepe et al. 2016), regions of
483 the genome harbouring extreme GEA signatures likely harbour many causal drivers of local
484 adaptation. However, given the covariation between structure and environment, it is difficult
485 to know what proportion of these regions are under selection vs. drift. Furthermore, repeated
486 GEA patterns between the latitudinal and elevational transects are non-independent, so it is
487 inappropriate to use statistical tests of whether repeated patterns exceed random
488 expectations. Designating any SNP with a GEA p-value in the bottom 1st percentile as a
489 "GEA outlier" and any SNP with excess ancestry in the genomic clines analysis as an
490 "ancestry outlier", we present results showing the overlap in outlier SNPs and genes
491 harbouring outliers, both among analyses and across the elevation and latitudinal datasets in
492 the Supplementary Material (supplementary figs. S29 and S30, Supplementary Material
493 online). It is more likely that the patterns of directional introgression are more evolutionarily
494 independent (see Discussion), so overlap of extreme signatures here is more informative. A
495 total of 18 genes were identified as candidates by all four analyses (both GEAs and both
496 cline analyses), exceeding the random expectation (expected 9.7 genes; $p = 0.0007$)
497 (supplementary Table S1 and supplementary fig. S29, Supplementary Material online).
498 Although this set of 18 candidate genes requires further study and validation, it comprises
499 several protein families with key roles in plant growth and environmental stress responses
500 (supplementary Text S3 and Table S1, Supplementary Material online), according to the
501 available annotation (Warren et al. 2015; Gagalova et al. 2022). More broadly, the less
502 stringent set of 638 candidates identified by either of our analyses includes several genes
503 previously reported to contribute to adaptation in other genomic studies of North American
504 conifers (supplementary Text S4 and Table S2, Supplementary Material online). Additionally,
505 we examined differences between the candidates identified in the latitudinal and elevational
506 analyses to highlight two distinct groups of SNPs: one putatively associated with elevational

but not latitudinal environmental pressures, and the other showing the opposite trend (supplementary Text S3, Supplementary Material online).

Discussion

Here, we assessed local adaptation, hybridization, and introgression patterns in two regions of the Western Canada interior spruce hybrid zone: a latitudinal gradient spanning ~1000 km of gradual environmental change and a narrower elevational contact zone spanning ~100km, including many regions with dramatic changes in environment over short spatial scales of < 10km. We observed dramatic changes in the proportion of white vs. Engelmann ancestry across different elevation/aspect sampling sites within several valley locations (Fig. 1c; Fig. 2). There are two main categories of explanation for the maintenance of highly differentiated populations across such small spatial scales: strong divergent natural selection and/or limited gene flow. Local adaptation between white and Engelmann spruce is well-established (De La Torre et al. 2013, 2014; Hornoy et al. 2015; Yeaman et al. 2016; Capblancq et al. 2022), but it is also possible that phenological barriers prevent mating between sites within locations, given that flowering time is known to respond plastically to temperature cues (Howe et al. 2003; De La Torre et al. 2013, 2014). Further study with transplant experiments and analysis of phenological barriers to reproduction would be needed to resolve this question. It is also unclear whether these repeated patterns of covariation between ancestry and environment represent a single colonization of all N-slopes by Engelmann and all S-slopes by white spruce, or initial colonization of the valleys followed by repeated expansion into the alpine areas by each species.

Our analyses revealed that patterns in the genomic basis of introgression are broadly similar between the elevational and latitudinal transects. We found strong repeatability in our F_{ST} and GEA analyses, with high correlation of the resulting F_{ST} and GEA p-values and a visually concordant distribution of peaks and valleys in the resulting Manhattan plots (Fig. 3; Fig. 4;

533 supplementary figs. S10 and S15, Supplementary Material online). While some repeatability
534 is expected given the non-independence involved when two hybrid zones arise from the
535 same parental species, the extent to which the patterns observed at broad latitudinal scales
536 (~1000 km) are repeated at the very narrow spatial scales of the elevational transect (<100
537 km) is remarkable (Fig. 3; Fig. 4). From these results alone, however, it is not possible to
538 ascertain whether highly divergent regions of the genome were driven by selection or drift,
539 but they do provide a set of candidate regions that likely include many of the causal drivers
540 of local adaptation, which has been well-studied at the phenotypic level (Jaramillo-Correa et
541 al. 2001; De La Torre et al. 2013; Liepe et al. 2016).

542 The picture becomes clearer with patterns of directional introgression, which are unlikely to
543 be repeated under a pure drift model if the populations in the middle of one transect are
544 evolving independently of those in the other. Our observation of extensive repeatability of
545 directional introgression (Figure 4d; supplementary fig. S28) is therefore more strongly
546 suggestive of a role for natural selection in shaping genomic patterns. However, it is not
547 possible from the data included here to conclusively test whether these populations are
548 evolving independently, as the dominant signal of relatedness is driven by ancestry
549 proportion, leaving little power to assess fine-scale patterns of gene flow. While average
550 dispersal distances in trees tend to be under 1 km per generation (Austerlitz et al. 2004),
551 both pollen and seed are known to travel long distances in conifers, which also exhibit high
552 degrees of outcrossing (O'Connell et al. 2007). It does seem unlikely, however, that gene
553 flow between populations in the middle of the two transects (most of which are separated by
554 > 100km; Fig. 1 and supplementary fig. S25, Supplementary Material online) could be high
555 enough to entrain their allele frequencies to drive repeated patterns, while also being limited
556 enough to allow the dramatic divergence we see within several locations in the elevation
557 transect (< 10km). Although further study is required to conclusively assess this evidence,
558 the most likely explanation for observed repeated patterns of directional introgression is
559 natural 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. S18, 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); however, not accounting for population structure is very likely to lead to elevated false-positive rates, making this a particularly challenging problem without a clear solution. When the goal is to identify a long list of candidates for future study, maximizing sensitivity while allowing for inclusion of false positives is preferable to using structure correction to strictly control the false positive rate and fail to detect genomic regions driving selection.

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 transect, there is less opportunity (fewer generations) 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. S19 and S20, Supplementary Material online). Given that common garden experiments show substantial evidence for local adaptation between the parental species, some of these regions may be shaped by strong divergent selection, which would further contribute to these patterns (Shchur et al. 2020). Unfortunately, the power of this analysis was limited by the restricted overlap among genomic regions targeted by the different sequence capture panels used for the two datasets. In addition, the fragmented state of the white spruce genome assembly and the limited resolution of the linkage map required us to rely on the rank order of SNPs within scaffolds anchored along the LGs of our super-scaffold assembly rather than on physical or recombinational (cM) distances, reducing

587 spatial precision. Therefore, further exploration of this hypothesis is warranted once a more
588 contiguous reference genome becomes available.

589 Our results also echo previous work suggesting that the interior spruce hybrid zone is
590 maintained by a bounded hybrid superiority model, where hybrids exhibit higher fitness than
591 parental species in intermediate environments (De La Torre et al. 2013, 2014). In our case,
592 this may be represented by high south-facing slopes in the middle of the elevational dataset
593 (locations 3, 4, 5, and 6; Fig. 1c), which combine features of the longer, dry summers that
594 favour *P. glauca* and the higher precipitation that favours *P. engelmannii*. Consistent with
595 this, south-facing slopes reported a higher frequency of balanced hybrids (hybrids close to
596 50:50 ratio), a balanced distribution of *P. glauca* dominant and *P. engelmannii* dominant
597 hybrids as well as lower frequencies of putative pure *P. glauca* individuals compared to
598 valleys, and putative pure *P. engelmannii* individuals compared to north-facing slopes (Fig. 2;
599 supplementary fig. S8, Supplementary Material online). The intermediate habitats provided
600 by south-facing slopes may allow hybrids to outperform both parental species by combining
601 traits such as faster growth with earlier bud setting and greater tolerance to snow load (De
602 La Torre et al. 2013, 2014).

603 Loci that repeatedly exhibit high differentiation, strong evidence of adaptation, and consistent
604 directional introgression constitute the strongest candidates for local adaptation and adaptive
605 introgression (Rieseberg et al. 1999; Buerkle and Lexer 2008; Gompert and Buerkle 2009;
606 Bierne et al. 2013; Simon et al. 2020; Westram et al. 2021; Langdon et al. 2022). The 18
607 strongest candidate adaptive loci identified here included several genes previously reported
608 for their central role in plant growth and environmental adaptation (supplementary Text S3
609 and supplementary Table S1, Supplementary Material online). Furthermore, several of the
610 genes present within our broader set of candidates (638 genes identified by either of our
611 analyses), were also reported to contribute to adaptation in similar studies of local adaptation
612 in other North American conifers (Hornoy et al. 2015; Yeaman et al. 2016; Depardieu et al.
613 2021; Capblancq et al. 2022) (supplementary Text S4 and Table S2, Supplementary Material

online). Additionally, differences in candidate loci between the elevational and latitudinal transects revealed several genes potentially involved in the response to environmental factors specific to latitude such as photoperiod and seasonality, as well as others potentially involved in regulating factors more specific to elevational changes, such as oxygen levels, UV radiation, temperature, snow and precipitation levels (supplementary Text S3, Supplementary Material online). However, we note that our candidates for local adaptation require further validation, and it is likely that this broader set of 638 genes includes false positives due to the difficulty in completely disentangling environmentally driven variation from neutral population structure effects. One way we tried to address this was through the analysis of genomic clines, as directional introgression replicated across hybrid zones characterized by different environmental contrasts would not be expected under a neutral drift model (Rieseberg et al. 1999; Buerkle and Lexer 2008; Gompert and Buerkle 2009; Bierne et al. 2013; Simon et al. 2020; Westram et al. 2021; Langdon et al. 2022). This is what we observed: directional introgression occurred largely in the same genomic regions and often in the same direction between the elevational and latitudinal transects, as reflected by the significant correlation of *BGC* alpha scores and overall highly concordant genome-wide patterns of directional introgression (Fig. 4; supplementary fig. S28, Supplementary Material online). This supports the idea that the 18 strongest candidate genes identified by all four analyses form a robust set of adaptive loci (supplementary Table S1 and fig. S29, Supplementary Material online), unlikely to be generated by drift.

Materials and Methods

Super-scaffold assembly construction

We built a super-scaffold assembly of the WS77111v2 white spruce (*P. glauca*) genome to use in subsequent analyses, by conservatively anchoring scaffolds to the latest publicly available *P. glauca* linkage map including 14,727 cDNAs (Gagalova et al. 2022). For this

640 process, we retrieved the sequences of an exhaustive set of 27,720 cDNAs, stemming from
641 the catalogue of spruce expressed genes assembled by Rigault et al. (2011), which we
642 filtered to retain only the sequences of the 14,727 cDNAs present in the *P. glauca* linkage
643 map (Gagalova et al. 2022). We then used *GMAP* (Wu et al. 2016) to align the retained
644 cDNAs sequences to the scaffolds of the WS77111v2 white spruce reference genome, and
645 the output was processed according to the following criteria: mapped cDNAs with coverage
646 above 50% and sequence identity above 90% were retained, while anything else was
647 discarded to ensure retention of only high confidence hits. Then, for each cDNA that satisfied
648 these mapping criteria, we extracted its best scaffold match based on the highest sequence
649 identity. In case a cDNA best mapping scaffold could not be determined as it mapped equally
650 well to multiple locations in the genome, it was discarded. This conservative approach
651 ensured the removal of difficult to resolve multi-mapped cDNAs. Finally, the scaffolds with
652 mapped cDNAs were ordered along the 12 spruce linkage groups according to the cDNAs
653 centimorgan (cM) positions in the linkage map.

654

655 **Sampling and sequencing**

656 We collected cambium tissue from 331 mature spruce individuals, including *P. glauca*, *P.*
657 *engelmannii* and their hybrids, from seven locations along the Trans-Canada highway going
658 through Canmore, Banff and Yoho in the summer of 2022 (Fig. 1c, elevational transect).
659 Each location included three distinct populations subjected to contrasting environmental
660 pressures: a north-facing high-altitude population, a south-facing high-altitude population,
661 and a valley population (Fig. 1c). Within each population, we sampled between 14 and 16
662 individuals (supplementary fig. S1, Supplementary Material online). Cambium tissue
663 samples were sent to Rapid Genomics LLC (Gainesville, Florida) for DNA extraction, library
664 preparation, and sequencing using a 10 Mb target approach.

665 We assembled a second dataset by downloading raw exome capture data for 575 mature
666 spruce individuals from across 249 populations distributed throughout western Canada (Fig.
667 1a, latitudinal transects), derived from previous studies (Suren et al. 2016; Yeaman et al.
668 2016). This data is publicly available in the Short Read Archive (SRP071805;
669 PRJNA251573).

670

671 **SNP calling**

672 We applied a uniform SNP calling pipeline to both datasets for consistency. First, *fastp*
673 (Chen et al. 2018) was used to remove sequencing adapters and for trimming raw paired
674 end FASTQ files, using default settings. Processed reads were then mapped to the latest
675 publicly available *P. glauca* reference genome (WS77111v2,
676 https://www.bcgsc.ca/downloads/btl/Spruce/) (Warren et al. 2015) using *BWA-mem* (Li and
677 Durbin 2009). Following mapping, we used *samtools* (Li et al. 2009) to convert the alignment
678 files generated from SAM to sorted BAM formats and discarded any alignment with mapping
679 quality below 10. *Picard* tools (available at https://broadinstitute.github.io/picard/) was then
680 used to remove potential PCR duplicates and to assign read groups, followed by indel
681 realignment with *GATK 3.8* (DePristo et al. 2011). Finally, SNP calling was performed using
682 *BCFtools mpileup* and *BCFtools call* (Danecek et al. 2021), based only on alignments with a
683 minimum mapping quality of 5 (−q 5). We filtered raw VCF files with *VCFtools* (Danecek et
684 al. 2011) to retain only biallelic SNPs with the following characteristics: genotyped in at least
685 70% of the individuals, quality above 30 (--minQ 30), individuals genotype quality above 20
686 (--minGQ 20), minimum read depth above 5 (--minDP 5) and minor allele frequency greater
687 than 0.05. Additionally, we filtered the VCF files of both datasets (latitudinal and elevational)
688 to retain only the SNPs located on the scaffolds that we were able to anchor to the available
689 linkage map (Gagalova et al. 2022) in our super-scaffold assembly construction.

690

691 **Population structure assessment**

692 We used *fastSTRUCTURE* (Raj et al. 2014) to assess population structure in both datasets,
693 using a simple prior and testing Ks ranging from 1 to 10, and then selected the K model that
694 maximizes the log-marginal likelihood lower bound (LLBO) of the data (K*ε). To determine
695 which species each *fastSTRUCTURE* cluster represented, we added whole genome
696 sequencing data of two pure species individuals cultivated at the arboretum of the Horsholm
697 University of Copenhagen (Denmark) and repeated the *fastSTRUCTURE* inference. This
698 data was retrieved from ENA under accessions SAMEA51442168 (*P. glauca*) and
699 ERX1862345 (*P. engelmannii*). Furthermore, we classified individuals as either pure *P.*
700 *glauca*, pure *P. engelmannii* or hybrid according to the value of the admixture coefficient (Q)
701 computed with *fastSTRUCTURE* at the chosen model complexity (K=2): $q \geq 0.95$ for either
702 pure species, while individuals with $0.05 < q < 0.95$ were classified as hybrids. Population
703 structure was further assessed using principal component analysis (PCA), implemented in
704 *Plink* version 2.0 (Chang et al. 2015). Marker-based realized genomic relatedness was
705 computed between all possible pairs of individuals according to the formula by VanRaden
706 (2008), implemented in the kin function of the R package *synbreed* (Wimmer et al. 2012).

707

708 **Population differentiation**

709 Weir and Cockerham F_{ST} (Weir and Cockerham 1984) was calculated on a per-SNP basis
710 among the 21 populations of the elevational dataset and the 249 populations of the
711 latitudinal dataset using the *VCFtools* function *weir-fst-pop* (Danecek et al. 2011), which
712 reports the average F_{ST} of all pairwise population comparisons for each SNP. Additionally,
713 F_{ST} between *P. glauca* and *P. engelmannii* was calculated within each dataset, selecting pure
714 species individuals based on the admixture coefficient computed with *fastSTRUCTURE* (q
715 ≥ 0.95). Using this assignment, nucleotide diversity π (Nei and Li 1979) was also computed

716 within each parental species (π) and between parental species (*dxy*) using piawka (Scott et
717 al. 2025).

718 Moran's I indices of spatial autocorrelation based on F_{ST} scores were computed using the R
719 package *moranfast* (https://github.com/mcooper/moranfast), using SNP position along both
720 scaffold and linkage group. To measure clustering of F_{ST} values along the genome, we
721 calculated Moran's I using SNP index positions, assigning each SNP a sequential number
722 (1, 2, 3, ...) based on its genomic order in our super-scaffold assembly. This approach treats
723 SNPs as equally spaced and captures how long similar values persist along scaffolds and,
724 more broadly, along linkage groups. It is intended to reflect the blockiness of the signal,
725 which may be related to haplotype structure. We used ordinal position of SNPs rather than
726 genomic positions because base-pair distances between scaffolds are unknown, due to the
727 highly fragmented reference genome. Additionally, estimation of Moran's I using genomic
728 position would be lower if a given set of SNPs were spread across long physical distances.
729 As it is expected that SNP density would be lower in regions with slow LD decay and longer
730 LD tracts, using genomic distances would erroneously bias the measurement downwards in
731 these regions. Ideally, this analysis would use recombination units (cM), but the available
732 linkage map lacks the resolution necessary to permit this.

733

734 **Genome environment association analysis**

735 We downloaded elevation data at 90-meter resolution from the NASA Shuttle Radar
736 Topography Mission (SRTM) (https://doi.org/10.5069/G9445JDF) and used it to compute
737 slope and aspect using the *terrain* function of the *raster* R package (Hijmans and Etten
738 2012). Additionally, we downloaded seven solar variables from the Global Solar Atlas
739 (https://globalsolaratlas.info): photovoltaic electricity output (PVOUT), global horizontal
740 irradiation (GHI), diffuse horizontal irradiation (DIF), direct normal irradiation (DNI), optimum
741 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-resolution data for the vast area covered by the latitudinal dataset, therefore we only included snowmelt day for the elevational dataset GEA, as we deemed it particularly well suited to discriminate between north-facing slopes, south-facing slopes and valleys (supplementary fig. S13, Supplementary Material online).

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

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

787

788 **Genomic clines analysis**

789 We performed a genomic clines analysis to identify SNP loci exhibiting excess ancestry from
790 either parental species relative to the average genomic background, as well as to identify
791 SNP loci with unusual rates of allelic-frequency transitions along the hybrid index. For this
792 analysis we used a Bayesian genomic clines model implemented in *BGC* (Gompert and
793 Buerkle 2011, 2012) and a logistic regression-based model implemented in *introgress*
794 (Gompert and Buerkle 2010). In *introgress*, we used both permutation and parametric tests

795 to compute p-values. We restricted each analysis to SNP loci with a frequency differential
796 between pure *P. glauca* and pure *P. engelmannii* of at least 0.6. Parental pure species and
797 hybrid populations were defined for both datasets according to the admixture coefficient
798 computed with *fastSTRUCTURE* (Q ≥ 0.95). This resulted in the following setups: 82 pure *P.*
799 *engelmannii*, 163 pure *P. glauca*, and 330 admixed individuals in the latitudinal dataset, and
800 99 pure *P. engelmannii*, 40 pure *P. glauca*, and 192 admixed individuals in the elevational
801 dataset. Furthermore, we conducted two additional *BGC* analyses on the latitudinal dataset,
802 using the same SNPs but down sampling the data to obtain a similar number of individuals
803 and parental population composition as in the elevational dataset, to allow for better
804 comparison. Subset 1 included 70 pure *P. engelmannii*, 86 pure *P. glauca*, and 200 admixed
805 individuals, while Subset 2 included 70 pure *P. engelmannii*, 45 pure *P. glauca*, and 194
806 admixed individuals. In our subsampling strategy, we prioritized removing individuals from
807 the latitudinal dataset which were located near the area sampled in the elevational dataset,
808 as well as populations from western British Columbia (supplementary fig. S25,
809 Supplementary Material online), to avoid possible signals of admixture with Sitka spruce
810 (Hamilton et al. 2012). In *BGC* we used 100,000 MCMC generations with 50,000 burn-in and
811 retained every 24th iteration. After assessing convergence, SNPs showing excess ancestry
812 or unusual rates of allelic-frequency transitions along the hybrid population were identified as
813 those for which the 99% equal-tail probability interval (ETPI) for the posterior probability
814 distribution of the clines parameters α or β did not overlap zero. In *introgress*, we used
815 default settings and ran both the parametric and permutation tests (Gompert and Buerkle
816 2010). Both tools were run within the *ClineHelpR* R package (Martin et al. 2021).

817

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

819 We used bedtools (Quinlan and Hall 2010) to identify the genes intersected by our top
820 candidate SNPs from the GEA and genomic clines analyses. We expanded the gene
821 sequences by adding 5000 bp flanks on either side, to include possible regulatory elements.

822 We used the gene models provided for the WS77111v2 *P. glauca* reference genome
823 (https://www.bcgsc.ca/downloads/btl/Spruce/). We used *Orthofinder2* (Emms and Kelly
824 2019) on each species proteome to identify orthology between white and red spruce, and
825 between white spruce and lodgepole pine. For white spruce (*P. glauca*), we used the
826 proteome for WS77111v2, available at
827 https://www.bcgsc.ca/downloads/btl/Spruce/Pglauca_WS77111/genome_annotation/WS7711
828 1v2/. As red spruce was mapped to Norway spruce (*P. abies*) in the study by Capblancq et
829 al. (2022), we used the proteome available for *P. abies* version 1 reference genome
830 (https://treegenesdb.org/org/Picea-abies). For lodgepole pine (*Pinus contorta*), we first
831 retrieved the entire transcriptome analysed in Yeaman et al. (2016) and used TransDecoder
832 (https://github.com/TransDecoder/TransDecoder) to identify the longest open reading frames
833 for each transcript and then translate them into protein sequences. Finally, for each species
834 proteome we selected a single primary transcript per gene according to the longest isoform,
835 and then *Orthofinder2* was run using default settings and independently for white spruce
836 versus Norway spruce and white spruce versus lodgepole pine.

837

838 ## Data Availability

839 Sequence data are deposited in the Short Read Archive under BioProject IDs PRJNA251573
840 (latitudinal dataset) and PRJNA1294281 (elevational dataset). Additional resources are
841 available on Zenodo (https://doi.org/10.5281/zenodo.17474986), including the order of
842 anchored scaffolds in the linkage map used to arrange the VCFs according to our super-
843 scaffold assembly, VCF files for the elevational and latitudinal datasets, and metadata with
844 individual coordinates and populations information.

845

846 ## Acknowledgements

We would like to thank Sally Aitken for many conversations that helped inspire this work, Ian MacLachlan, Jason Mogilefsky, Vashti Dunham, and Tim McCready for help obtaining funding and overall planning of the project, and staff at the Canadian and Alberta parks systems for help with securing permits. Thanks also to Michael Williamson for help 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.

References

1. Aguillon SM, Dodge TO, Preising GA, Schumer M. 2022. Introgression. *Current Biology* 32.
2. Anderson E. 1953. Introgressive hybridization. *Biological Reviews* 28:280–307.
3. Austerlitz F, Dick CW, Dutech C, Klein EK, Oddou-Muratorio S, Smouse PE, Sork VL. 2004. Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology* 13:937–954.
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. Booker T, Yeaman S, Whitlock M. 2023. The WZA: A window-based method for characterizing genotype-environment association. *Mol. Ecol. Resour.* 24(2):e13768.
12. 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.
13. Buerkle CA, Lexer C. 2008. Admixture as the basis for genetic mapping. *Trends in Ecology & Evolution* 23:686–694.
14. Capblancq T, Forester BR. 2021. Redundancy analysis: A Swiss Army Knife for Landscape Genomics. *Methods in Ecology and Evolution* 12:2298–2309.
15. 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.
16. 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.
17. Chaturvedi S, Lucas LK, Buerkle CA, Fordyce JA, Forister ML, Nice CC, Gompert Z. 2020. Recent hybrids recapitulate ancient hybrid outcomes. *Nature Communications* 11.
18. Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:884–890.
19. Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* 23:3133–3157.
20. 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.
21. 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.
22. 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.
23. 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.
24. 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.
25. 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.
26. Emms DM, Kelly S. 2019. Orthofinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology* 20.
27. 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.
28. 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.
29. 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.
30. Gompert Z, Buerkle CA. 2010. introgress: A software package for mapping components of isolation in hybrids. *Mol Ecol Resour*. 10:378–384.
31. Gompert Z, Buerkle CA. 2011. Bayesian estimation of Genomic Clines. *Molecular Ecology* 20:2111–2127.
32. Gompert Z, Buerkle CA. 2012. BGC: Software for bayesian estimation of Genomic Clines. *Mol Ecol Resour*. 12:1168–1176.
33. 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.
34. 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.
35. Hamilton JA, Miller JM. 2015. Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conservation Biology* 30:33–41.
36. Harrison RG, Larson EL. 2014. Hybridization, introgression, and the nature of Species Boundaries. *Journal of Heredity* 105:795-809.
37. 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.
38. 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.
39. Hijmans RJ. 2010. Raster: Geographic Data Analysis and modeling. CRAN: Contributed Packages.
40. 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.
41. Horton, KW 1959. Characteristics of subalpine spruce in Alberta. Canada Department of Northern Affairs, Natural Resources, Forest Research Division, *Technical Note* 76.
42. Howe GT, Aitken SN, Neale DB, Jermstad KD, Wheeler NC, Chen TH. 2003. From genotype to phenotype: Unraveling the complexities of cold adaptation in Forest Trees. *Canadian Journal of Botany* 81:1247–1266.
43. 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.
44. 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.
45. 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.
46. 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.

47. 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.
48. 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.
49. Levin DA, Francisco-Ortega J, Jansen RK. 1996. Hybridization and the extinction of rare plant species. *Conservation Biology* 10:10–16.
50. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760.
51. 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.
52. 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.
53. 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.
54. Lotterhos KE, Whitlock MC. 2014. Evaluation of demographic history and neutral parameterization on the performance of fst outlier tests. *Molecular Ecology* 23:2178–2192.
55. 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.
56. Lotterhos KE. 2019. The effect of neutral recombination variation on genome scans for selection. G3 Genes|Genomes|Genetics 9:1851–1867.
57. Mallet J. 2005. Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* 20:229–237.
58. Martin BT, Chafin TK, Douglas MR, Douglas ME. 2021. ClineHelpR: An R package for genomic cline outlier detection and visualization. *BMC Bioinformatics* 22.
59. 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.
60. Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences* 76:5269–5273.
61. O'Connell LM, Mosseler A, Rajora OP. 2007. Extensive long-distance pollen dispersal in a fragmented landscape maintains genetic diversity in white spruce. *Journal of Heredity* 98:640–645.
62. 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.
63. Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
64. Raj A, Stephens M, Pritchard JK. 2014. FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* 197:573–589.
65. 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.

66. 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.
67. 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.
68. 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.
69. Scott AD, Kolesnikova UK, Glushkevich A, Steinmann L, Tikhomirov NP, Pfordt U, Bohutínská M, Burns R, Seregin AP, Kolar F, et al. 2025. Multiple autopolyploid arabidopsis lyrata populations stabilized by long-range adaptive introgression across Eurasia. *Molecular Biology and Evolution* 42.
70. Sedghifar A, Saelao P, Begun DJ. 2016. Genomic patterns of geographic differentiation in drosophila simulans. *Genetics* 202:1229–1240.
71. 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.
72. 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.
73. 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.
74. 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.
75. Taylor, TMC. 1959. The taxonomic relationship between Picea glauca (Moench) Voss and Picea engelmannii Parry. *Madrono* 15: 111–115.
76. VanRaden PM. 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91:4414–4423.
77. 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.
78. 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.
79. Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of Population Structure. *Evolution* 38:1358.
80. 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.
81. 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.
82. 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.

83. 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.
84. 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.

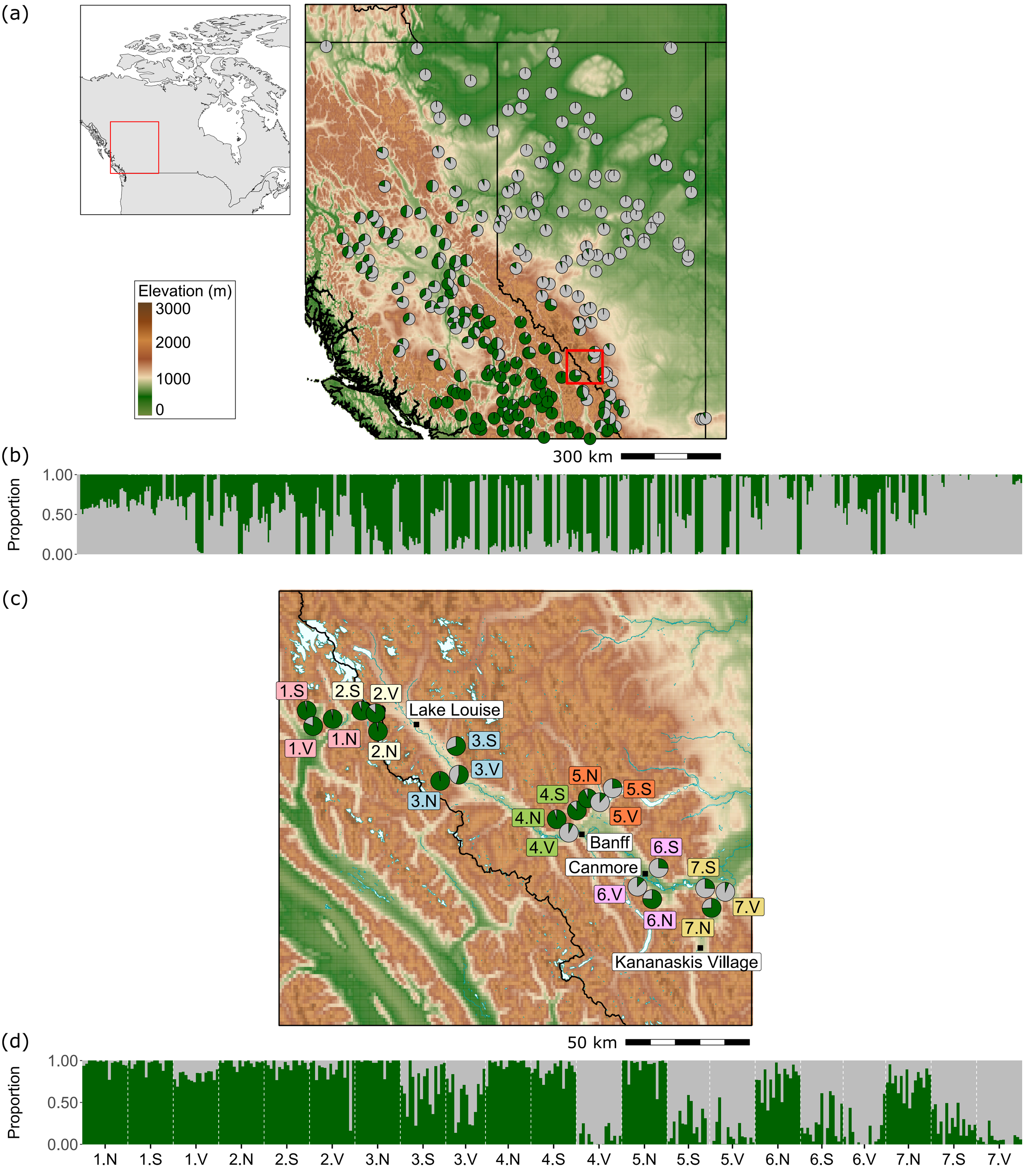
Figures

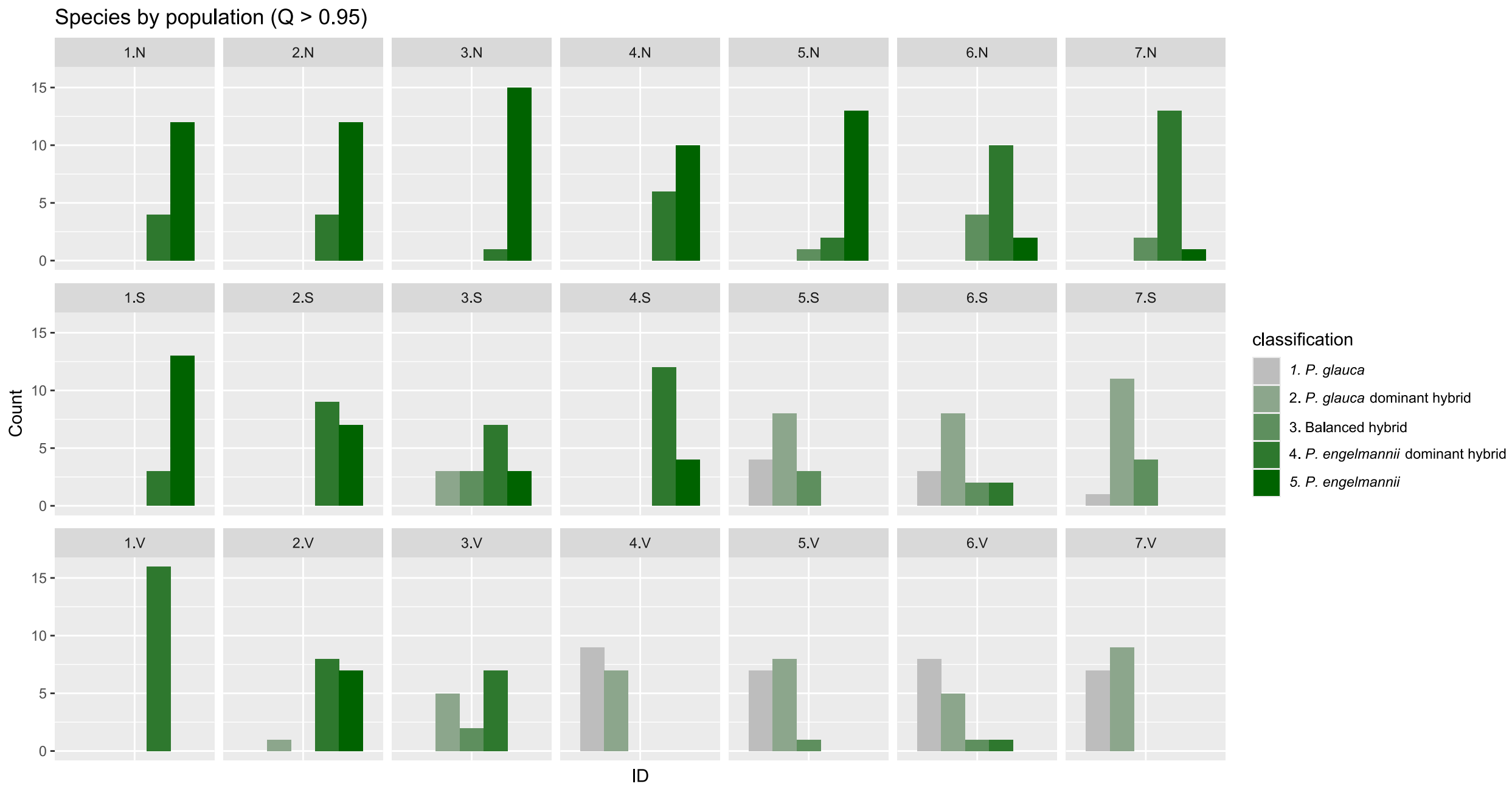
Fig. 1. Geographic locations of populations from the (a) latitudinal and (c) elevational datasets in Canada. Pie size reflects the number of individuals per population. Grey indicates *P. glauca* ancestry; green indicates *P. engelmannii* ancestry. The small overview map in (a) shows the location of the latitudinal transect samples in Canada. The red square in (a) marks the location of the elevational transect samples 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.

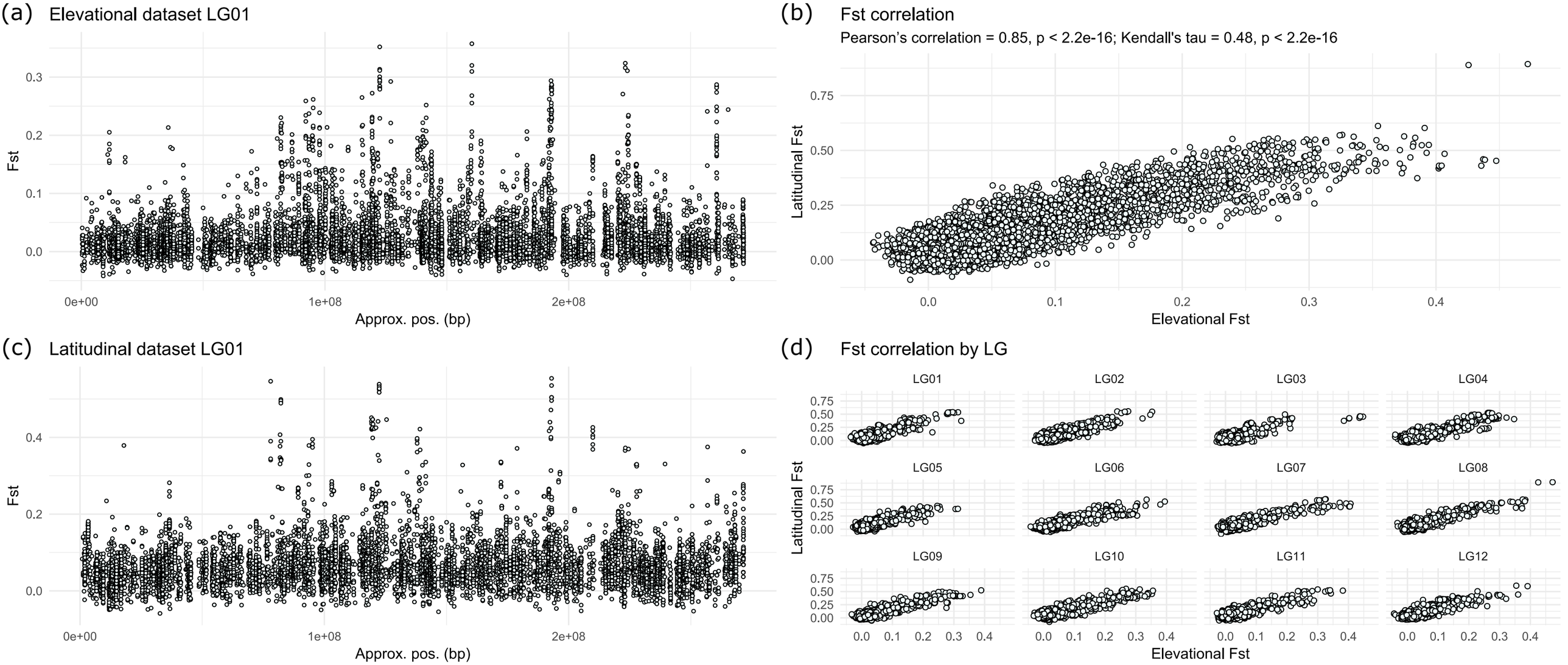
Fig. 2. Distribution of individual's ancestry category across locations and environments in the elevational datasets (N: north-facing slopes, S: south-facing slopes, V: valleys). Individuals were classified by ancestry according to the admixture coefficient computed with *fastSTRUCTURE*, using the following 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* ($q < 0.05$).

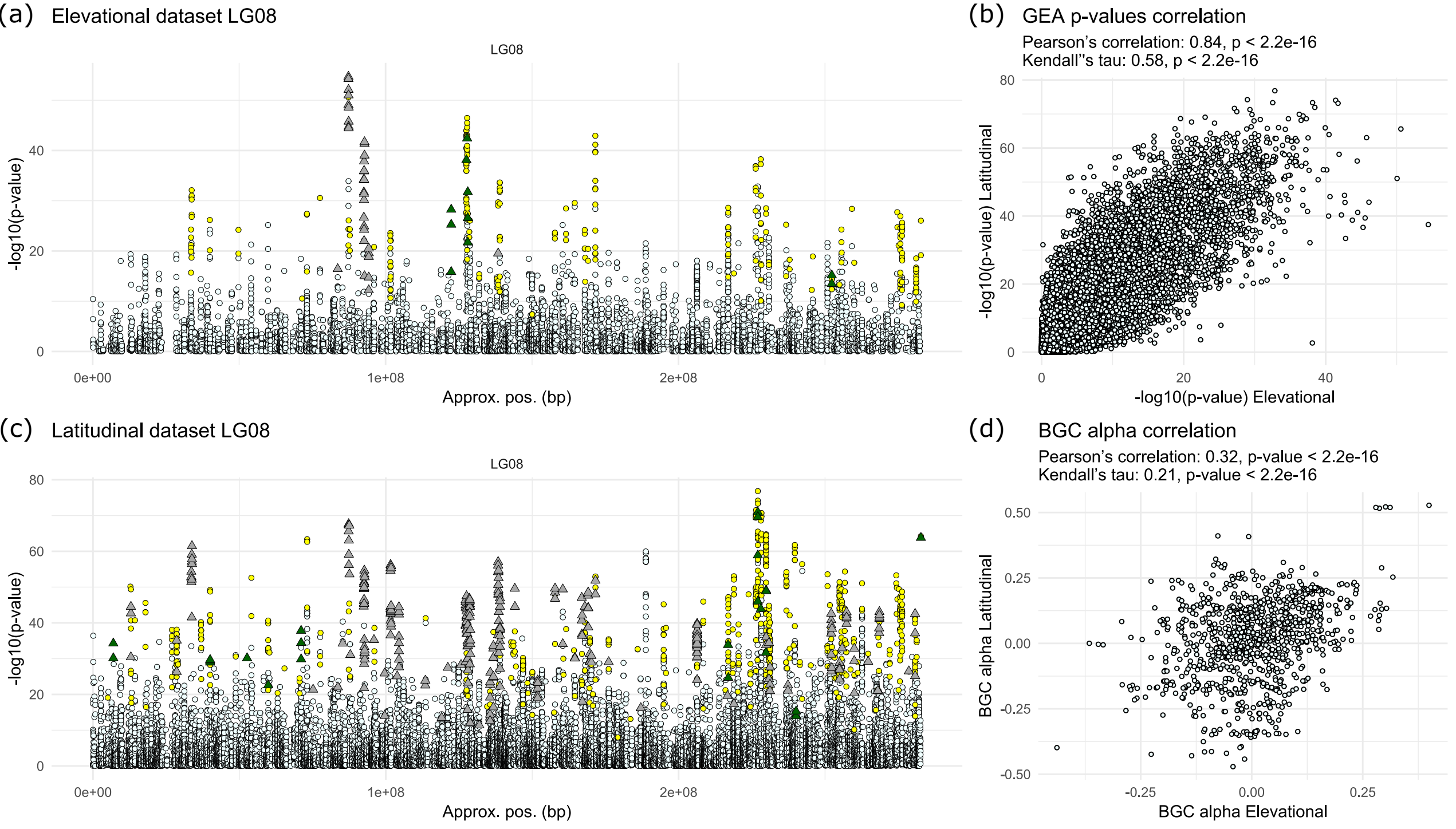
Fig. 3. Genomic patterns of among population divergence (F_{ST}). Manhattan plots of F_{ST} values along LG01 in the (a) elevational and (c) latitudinal dataset. (b) Genome-wide correlation in F_{ST} between the shared SNPs of the elevational and latitudinal datasets. (d) Correlation in F_{ST} by linkage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

Fig. 4. Genomic patterns of Genotype-Environment Association (GEA) and genomic clines. Manhattan plots of GEA p-values along LG08 in the (a) elevational and (c) latitudinal dataset. SNPs showing ancestry excess in the genomic clines (*BGC*) analyses are represented by triangles, coloured as follows: dark green indicates *P. engelmannii* ancestry excess, and grey indicates *P. glauca* ancestry excess. Yellow circles represent SNPs tested in the *BGC* analysis (parental frequency differential > 0.6) but not showing ancestry excess. Small pale-azure circles denote SNPs included only in the GEA analysis but not in the *BGC* analysis. (b) Genome-wide correlation of GEA p-values between the shared SNPs of the elevational and latitudinal datasets. (d) Genome-wide correlation of *BGC* α values between the shared SNPs of the elevational and latitudinal datasets. Note that the significance of p-values in panels (b) and (d) is overestimated as an assumption of the test is violated due to non-independence among linked SNPs.


(a)
Elevation (m)
3000
2000
1000
0
300 km
(b)
Proportion
1.00
0.50
0.00
(c)
1.S
2.S
2.V
Lake Louise
1.N
1.V
2.N
3.S
3.V
3.N
5.N
5.S
4.S
5.V
4.N
Banff
4.V
6.S
Canmore
7.S
6.V
7.V
6.N
7.N
Kananaskis Village
50 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
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: 0.84, p < 2.2e-16
Kendall''s tau: 0.58, 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