This document is the accepted manuscript version of the following article:
Dauphin, B., Wüest, R. O., Brodbeck, S., Zoller, S., Fischer, M. C., Holderegger, R., … Rellstab, C. (2020). Disentangling the effects of geographic peripherality and habitat suitability on neutral and adaptive genetic variation in Swiss stone pine. Molecular Ecology, 29(11), 1972-1989. https://doi.org/10.1111/mec.15467

Disentangling the effects of geographic peripherality and habitat suitability on neutral and adaptive genetic variation in Swiss stone pine

Benjamin Dauphin[1*], Rafael O. Wüest[1], Sabine Brodbeck[1], Stefan Zoller[2], Martin C. Fischer[3], Rolf Holderegger[1,3], Felix Gugerli[1], & Christian Rellstab[1*]

[1]WSL Swiss Federal Research Institute, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland

[2]Genetic Diversity Centre (GDC), ETH Zurich, 8092 Zurich, Switzerland

[3]Institute of Integrative Biology (IBZ), ETH Zurich, 8092 Zurich, Switzerland

*Authors for correspondence: benjamin.dauphin@bluewin.ch, christian.rellstab@wsl.ch

Running head: Drivers of genetic diversity in *P. cembra*

Abstract

It is generally accepted that the spatial distribution of neutral genetic diversity within a species' native range mostly depends on effective population size, demographic history, and geographic position. However, it is unclear how genetic diversity at adaptive loci correlates with geographic peripherality or with habitat suitability within the ecological niche. Using exome-wide genomic data and distribution maps of the Alpine range, we first tested whether geographic peripherality correlates with four measures of population genetic diversity at >17,000 SNP loci in 24 Alpine populations (480 individuals) of Swiss stone pine (*Pinus cembra*) from Switzerland. To distinguish between neutral and adaptive SNP sets, we used four approaches (two gene diversity estimates, F_{ST} outlier test, and environmental association analysis) that search for signatures of selection. Second, we established ecological niche models for *P. cembra* in the study range and investigated how habitat suitability correlates with genetic diversity at neutral and adaptive loci. All estimates of neutral genetic diversity decreased with geographic peripherality, but were uncorrelated with habitat suitability. However, heterozygosity (H_e) at adaptive loci based on Tajima's *D* declined significantly with increasingly suitable conditions. No other diversity estimates at adaptive loci were correlated with habitat suitability. Our findings suggest that populations at the edge of a species' geographic distribution harbour limited neutral genetic diversity due to demographic properties. Moreover, we argue that populations from suitable habitats went through strong selection processes, are thus well adapted to local conditions, and therefore exhibit reduced genetic diversity at adaptive loci compared to populations at niche margins.

Keywords

conifers, exome capture, gene diversity, geographic peripherality, habitat suitability, *Pinus cembra*

Introduction

Mutations as a source of genetic diversity are the major driving force of evolution (Nei, 2013). Most of the new variants behave neutrally, i.e., are not subject to natural selection (Nei, Suzuki, & Nozawa, 2010), but those variants in the small part of the genome that is shaped by selection allow populations to adapt to environmental change (Aitken, Yeaman, Holliday, Wang, & Curtis-McLane, 2008). This adaptive capacity or evolvability of populations is often based on standing genetic variation inherited through generations rather than on new mutations spontaneously arising in populations (Barrett & Schluter, 2008; Houle, 1992). This is particularly true for species with a long generation time (e.g. hydrozoans, reptiles, sponges, ferns, trees), for which the restricted temporal scale of current rapid environmental change leaves little chance for new beneficial mutations to spread across a population. Both standing genetic variation and recent mutations may confer adaptive capacity to novel environmental conditions. Besides mutations, gene flow among populations is also a key process for introducing new alleles into a population (Slatkin, 1985).

Aside from contemporary biotic or abiotic constraints, the geographic distribution of species primarily results from their demographic history, e.g., following contraction or expansion cycles due to glacial oscillations. The central abundance hypothesis (CAH; Brown, 1984; Hengelveld & Haeck, 1982) assumes that species abundance is largest at the centre of its geographical range and decreases gradually towards peripheral areas. However, empirical studies have shown that a large number of species do not conform to the CAH (Sagarin & Gaines, 2002). This is likely due to the fact that the geographic position within a species' range (i.e. geographic peripherality, GP, or centrality) and habitat suitability (HS) are often not spatially correlated and a species' occurrence is mostly driven by habitat conditions. In this context, several descriptors were proposed to characterise species presence based on

ecological niche conditions (Martínez-Meyer, Díaz-Porras, Peterson, & Yáñez-Arenas, 2013). In the HS approach of Martínez-Meyer et al. (2013), populations that are close to the niche centre (optimal habitat conditions) are considered core populations, whereas those distant from the niche centre are considered marginal populations. In cases where geographic and environmental features of a species' habitat are uncorrelated, using both concepts of GP and HS allows disentangling the geographic and environmental determinants of population dynamics.

The CAH can conceptually be applied in the context of genetic diversity, because large central populations likely have large census and effective population sizes (N_c and N_e) as well as high among-population gene flow, resulting in high genetic diversity (Macdonald, Llewelyn, Moritz, & Phillips, 2017). Therefore, past demographic history, in concert with selection processes, strongly shape patterns of genetic variation across a species' range (Felsenstein, 1976). Several empirical studies have shown that overall genetic diversity of populations is related to their geographic position within the species' range, with central populations harbouring higher genetic diversity compared to peripheral populations (Eckert, Samis, & Lougheed, 2008; Lee-Yaw, Fracassetti, & Willi, 2018; Lira-Noriega & Manthey, 2014). Nevertheless, this pattern seems difficult to generalise. Analyses of closely related species (e.g. *Bombus* spp. or *Cardellina* spp.; Lira-Noriega & Manthey, 2014) have shown contrasting and species-dependent patterns. The mentioned study additionally used a niche centroid-based approach to correlate the distance of populations from the niche centre with their genetic diversity and found, for some species, that core populations harboured highest genetic diversity, with a gradual decrease in genetic diversity towards marginal populations. This pattern is expected when HS is negatively correlated with GP.

The findings described above characterised overall (including genome-wide) genetic diversity, without differentiating between neutral and adaptive loci. Only a small fraction of the genome is supposedly shaped directly by natural selection (Exposito-Alonso, Burbano, Bossdorf, Nielsen, & Weigel, 2019; Fischer, Foll, Heckel, & Excoffier, 2014; Shapiro et al., 2007). Therefore, overall genetic diversity basically represents neutral genetic diversity shaped by neutral processes such as demographic history, genetic drift, and gene flow, which is largely uninformative regarding adaptive processes. In contrast, loci under selection often show distinct patterns of allele frequencies compared to those that evolve neutrally (Savolainen, Lascoux, & Merilä, 2013). In past investigations of GP and HS, little attention has been given to such a partitioning of genetic diversity at neutral and adaptive loci, despite high interest in conservation and population genetics. Investigations on the evolutionary and environmental drivers of genetic variation have often been based on overall or even solely neutral genetic diversity (e.g. Lei, Wang, Liu, He, & Li, 2015; Šurinová, Hadincová, Vandvik, & Münzbergová, 2019). There is thus a clear need to separately analyse neutral and adaptive regions in the genome to disentangle the geographic and environmental drivers of genetic variation. Moreover, it is still an open question how genetic diversity at adaptive loci varies across a species' range, and to what extent habitat suitability affects this diversity.

Generally, it is assumed that peripheral populations occur in less suitable and less stable habitats, and often in restricted and small habitat patches. Hence, sizes of peripheral and marginal populations are likely small, and populations might be rather young, except if they are relicts (e.g. refugial populations during the last glacial maximum). Moreover, colonisation of peripheral sites may have originated from different sources, and they are often far from each other. As a consequence, low gene flow among peripheral populations is expected, and genetic drift might be strong. Taken together, this situation is expected to lead

to low neutral genetic diversity in peripheral and in marginal populations (Figure 1a, b; Hampe & Petit, 2005). In contrast, populations in central geographic positions or at the core of a species' niche are generally established in suitable habitat patches. Populations are thus dense, numerous, and can be old. Consequently, neutral genetic diversity should be high, and gene flow is expected to homogenise allele frequencies at neutral loci.

Patterns of genetic diversity at adaptive loci are primarily determined by two main factors (Felsenstein, 1976; Slatkin, 1973); the strength of selection (selection coefficient [s]), itself influenced by effective population size (N_e; Gravel, 2016), and the counteracting effect of gene flow (i.e. migration rate [m]). In the scenario where m > s, selection might not be effective in small marginal populations, especially in unstable habitats where selective pressure constantly varies. Large core populations should experience high gene flow that leads to high genetic diversity also at adaptive loci; these populations should therefore be more diverse than marginal populations as a result of their large N_e (dashed line in Figure 1d). In the alternative scenario (m < s, solid line in Figure 1d), strong selection generally leads to low genetic diversity at adaptive loci. Large core populations would have adapted to their highly suitable habitat, which might result in low genetic diversity at adaptive loci. This is, however, only the case for single populations; overall genetic diversity at adaptive loci across all core populations can nevertheless be high, in particular if single populations show different genetic mechanisms to adapt to a similar habitat and, hence, selective pressures (Rellstab et al., 2017). Selection in populations of low habitat suitability might have been less efficient in pruning mal-adapted alleles, leading to populations with higher genetic diversity at adaptive loci than in populations occurring in highly suitable habitat. In both scenarios on the relative importance of m *versus* s, we hypothesise to find no correlation between GP and genetic diversity at adaptive loci, because selection is exerted by the environment and not by

132 neutral geographic processes (Figure 1c). However, since N_e is affecting the effectiveness of
133 selection and is hypothesised to be correlated to GP (see above), we acknowledge that another
134 possible scenario suggests that genetic diversity at adaptive loci is influenced indirectly by N_e,
135 potentially leading to a similar pattern as in Figure 1a.
136 To test the hypotheses presented in Figure 1, one ideally considers a study species that
137 (a) experiences strong natural selection, (b) occupies heterogeneous habitats, and (c) exhibits
138 no strong correlation between GP and HS in order to disentangle these two components that
139 often covary in space. Under these premises, Alpine Swiss stone pine (*Pinus cembra*)
140 provides an ideal study system, because it grows at the timberline ecotone that exhibits high
141 selection pressures (e.g. by frost and high UV radiation) and occurs in heterogeneous habitats
142 including various micro-topographic conditions (e.g. slope, exposure, and drainage). The
143 species is also known to have experienced a complex re-colonisation history after the last
144 glacial maximum with putatively multiple refugia in the periphery of the Alps (Höhn et al.,
145 2009). As a consequence, GP and HS are decoupled in this system to a certain extent,
146 allowing us to assess their relative effects on genetic diversity. Here, we use presence/absence
147 data, species distribution modelling, and exome-wide genomic data of 480 trees in 24
148 populations to investigate the relationships between GP/HS and genetic diversity at neutral
149 and adaptive loci. We show how patterns of genetic diversity are distributed across space and
150 how genetic diversity is influenced by geographic position and environmental conditions. We
151 further demonstrate how important it is to distinguish neutral and adaptive loci to fully
152 account for the nature of genetic diversity and its respective drivers.
153

Materials and Methods

Study species and area

Swiss stone pine (*Pinus cembra* L.) is a five-needle, closed-cone pine of subgenus *Strobus* (Gernandt, Geada López, Ortiz García, & Liston, 2005). It has a restricted geographical range in the Central European Alps and the Carpathian Mountains and is found at the upper range of forested area (1,500–2,400 m a.s.l.) up to the colonisation front at the tree line. It is a keystone species of the timberline ecotone that has experienced substantial population decline over the last two centuries, mainly as a consequence of human activity such as forest clearing for pastures and ungulate grazing (Höhn et al., 2009; Motta & Nola, 2001). The species is mostly outcrossing and shows high levels of gene flow supported by wind pollination (Salzer & Gugerli, 2012). However, dispersal by seed, primarily through spotted nutcracker (*Nucifraga caryocatactes*), is spatially limited (Salzer, 2011). Other biotic drivers (e.g., understory vegetation) and climatic factors seem to play an important role in post-dispersal recruitment (Meier et al., 2010; Neuschulz, Merges, Bollmann, Gugerli, & Böhning-Gaese, 2018). As most conifers, *P. cembra* has a complex and very large genome (29.3 Gbp, 2n = 24; Zonneveld, 2012). Switzerland, with its long and steep environmental gradients, offers a unique opportunity to study environmental marginality for an alpine species such as *P. cembra* (Figure S1).

Sampling and collection of occurrence data

We sampled 24 populations across a large environmental gradient covering the two main phylogeographical lineages of *P. cembra* (Gugerli, Rüegg, & Vendramin, 2009) in the Swiss Alpine range (Table S1, Figure 2). In each population, we sampled 20 georeferenced juvenile trees. Tree ages were estimated in the field by counting annual shoot increments, targeting

178 juveniles aged about 10-20 years. In total, we collected needle samples of 480 individuals for
179 molecular analyses.
180 We obtained species occurrence data from the fourth Swiss National Forest Inventory
181 (NFI4 recorded in the years 2009-2017; Fischer & Traub, 2019) and from InfoFlora, the Swiss
182 national floristic database (https://www.infoflora.ch). The data were manually curated by
183 removing non-native occurrences (e.g., possibly planted trees), non-validated occurrences
184 (e.g., uncertain species identification), or records with imprecise geographical coordinates
185 (precision >50 m). After this filtering, we retained 1,876 presence observations (1,621 from
186 InfoFlora, 255 from NFI4) and 6,057 absence records from NFI4 (Figure 2).
187
188 *Environmental data*
189 We collected topographic and climatic data to characterise environmental conditions in each
190 population to (a) carry out species distribution modelling (SDMs, Guisan & Zimmermann,
191 2000) and (b) correlate environmental variation with genomic variation in environmental
192 association analysis (EAA, Rellstab, Gugerli, Eckert, Hancock, & Holderegger, 2015). For
193 topography, we used a 100 m digital elevation model (aggregated from the DHM25 at 25 m
194 resolution; Swisstopo, 2004) to derive 15 variables (Table S2) based on their informative
195 power at local scale (Leempoel et al., 2015). We calculated morphometric, hydrologic, and
196 radiation grids for Switzerland using SAGA 6.2 (details in Table S3; Conrad et al., 2015).
197 Climatic data consisted of 19 bioclimatic predictors (as described at http://chelsa-
198 climate.org/bioclim/), which were calculated using monthly aggregated temperature and
199 precipitation data for the reference period 1981–2010. The monthly data were based on
200 weather station data from the Federal Office of Meteorology and Climatology MeteoSwiss

interpolated to a resolution of 100 m × 100 m using the 100 m digital elevation model and the DAYMET software (Thornton, Running, & White, 1997).

Geographic peripherality and habitat suitability

To characterise the GP of each population, we used the geographic distribution of the species in the Alps (Caudullo, Welk, & San-Miguel-Ayanz, 2017) to assess the Euclidian distance of the centroid of each sampled population to the closest range limit. This distance was then converted to a continuous peripherality index:

$$GP_i = (D_{max} - D_i) + 1 \ ,$$

where GP_i is the geographic peripherality of population i, D_{max} is the maximum Euclidian distance [km] to the closest range limit across all populations, and D_i is the Euclidian distance of population i to its closest range limit. The fixed added term (+1) avoids a null value for the population that has the highest Euclidian distance to the closest range limit. High GP values indicate that a population is close to the species' range limit (i.e. is a peripheral population), and low GP values represent central populations. Note that the GP index is limited to the scale of the European Alps, ignoring the fragmented distribution of the species in the Carpathian Mountains.

We used a species distribution modelling (SDM) approach to characterise the distribution of suitable habitat for *P. cembra* in Switzerland. Following current standards (Araújo et al., 2019), we constructed an ensemble of SDMs using the following five SDM algorithms and packages of the R statistical software (version 3.4.4; R Core Team, 2019): (1) generalised linear model (GLM; Nelder & Wedderburn, 1972; using STATS); (2) generalised additive model (GAM; Hastie & Tibshirani, 1990; using MGCV; Wood, 2011); (3) random forest (RF; Breiman, 2001; using RANDOMFOREST; Liaw & Wiener, 2002); (4) artificial

neural networks (ANN; Ripley, 1996; using NNET; Venables & Ripley, 2002); (5) maximum-entropy (MAXENT, Phillips, Aneja, Kang, & Arya, 2006; using DISMO; Hijmans, Phillips, Leathwick, Elith, & Hijmans, 2017). We fitted GLM using linear and quadratic terms and GAM with smooths of up to four degrees of freedom, while assuming binomial error distribution and logit link for both. RF and ANN were tuned, with resulting optimal parameters as follows: minimal terminal node size was set to three, number of trees to 1,000, and the number of candidate variables at each split to three for RF, whereas number of hidden layers was set to four and weight decay to 0.1 for ANN. We used default settings for MAXENT, except that we set the minimal number of observations for including hinge and product features to 100 and 150, respectively. All variables were standardised prior to model fitting.

Before fitting the SDMs, we applied a variable selection procedure that chooses the best performing predictors while simultaneously avoiding high collinearity. We followed the procedure described in Wüest et al. (2020) that first fits a logistic regression for each predictor including a linear and quadratic term and cross-validates these univariate models using repeated split-sample cross-validation (details follow the procedure for the evaluation of model performance outlined below). We averaged the true skill statistic (TSS; Allouche, Tsoar, & Kadmon, 2006) on the out-of-bag portion in each repetition to rank the predictors according to their predictive power. As a final step, we reduced the predictor set to only contain variables with pairwise Pearson correlations of $|r| < 0.7$, while giving preference to variables with high predictive power.

Model performance was assessed using cross-validation. We repeatedly split our data into 70% training and 30% testing data. In each repeat, we fitted the five SDM algorithms to the training data, and transformed the predicted probabilities of occurrence for the testing data

into binary presence and absence using a threshold that optimises TSS. This TSS was then calculated for all five models and each of the 100 repeats to serve as a measure of the model's predictive performance.

We generated ensemble predictions of habitat suitability for each of the sampled and georeferenced trees as follows. Using the relevant environmental predictors extracted for each of the individual tree locations, we predicted the probability of occurrence using the five fitted models. The ensemble consisted of a weighted average, for which we used model-specific TSS values (averaged over the 100 cross-validation repeats) as weights. These ensemble predictions at the level of individual trees were then averaged to obtain population-level habitat suitability. We further generated spatial projections of all single SDM algorithms as well as the weighted ensemble and standard deviation (among the five SDM algorithms) across Switzerland for illustrative purposes.

Variable importance of each variable across all five SDM algorithms was assessed by repeatedly permuting the values of a predictor variable (only one variable at a time) and predicting the probability of occurrence using a permuted dataset. These predictions p_{shuff} were then compared to the original predictions (p_{ref}; no permutation of any predictor variable) to generate an importance measure defined as $1 - \rho_{p_{shuff},p_{ref}}$ (where $\rho_{p_{shuff},p_{ref}}$ is the correlation between p_{shuff} and p_{ref}). This importance was calculated for each model and variable in each of the repeats. To facilitate interpretation, we scaled variable importance averaged across the repeats to sum to 100%.

Exome capture sequencing, SNP calling, and filtering steps

We carried out DNA extraction, library preparation, and exome capture as described in Rellstab et al. (2019). Briefly, high-quality DNA of 20 trees per population was used to

produce equimolar DNA pools for all 24 populations for pooled sequencing (Pool-Seq; Rellstab, Zoller, Tedder, Gugerli, & Fischer, 2013; Schlötterer, Tobler, Kofler, & Nolte, 2014), which has shown to yield accurate estimates of allele frequencies in this sequencing approach (Rellstab et al., 2019). We generated barcoded libraries (average insert size of 550 bp) using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Massachusetts, USA) and subsequently performed probe hybridisation using the MYcroarray myBaits Custom Capture Kit. The 24 hybridised libraries were then sequenced on four lanes of an Illumina HiSeq 4000 (paired-end reads of 150 bp) at the Functional Genomics Center Zurich (FGCZ, Zurich, Switzerland) and Fasteris (Geneva, Switzerland; Table S4).

Following Rellstab et al. (2019), we trimmed and filtered raw reads with TRIMMOMATIC 0.35 (Bolger, Lohse, & Usadel, 2014) using a quality threshold of 20 on both forward and reverse reads. We then mapped the remaining reads back to those transcripts of the reference transcriptome that contained probe bases using BOWTIE 2.3.0 (Langmead, Trapnell, Pop, & Salzberg, 2009), and performed variant (i.e. SNP) and invariant site calling using GATK 3.8 (McKenna et al., 2010) with ploidy set to 40 (i.e. number of chromosomes sequenced per pool of 20 diploid individuals), a coverage ≥40×, and a mapping quality/depth ratio ≥0.25. To get rid of putatively paralogous genes, variant and invariant calling was carried out only for the 4,950 single-copy contigs as determined in Rellstab et al. (2019). These authors used HDPLOT (McKinney, Waples, Seeb, & Seeb, 2017) to exclude putatively paralogous contigs based on excess heterozygosity and deviation from usual allele balance (read ratio). To conduct population genetic analyses, we assembled a SNP set based on two additional filters to exclude weakly supported SNPs: excluding SNPs with (i) a minor allele frequency (MAF) ≤2.5% across populations (i.e. one chromosome in a pool) and (ii) missing

296 data in at least one population. We used the resulting SNP set in the format of either
297 population allele frequency or read count data for downstream analyses (see below).
298
299 *Population genetic structure and diversity*
300 To investigate population genetic structure, we performed a principal component analysis
301 (PCA) using allele frequencies of the complete SNP set with the *prcomp* function from the R
302 package STATS (centring and scaling by default; R Core Team, 2019). Based on read count
303 data, we carried out a hierarchical clustering analysis from the dissimilarity matrix Ω ($d_{ij} =$
304 $1 - \rho_{ij}$) generated with BAYPASS 2.1 (Gautier, 2015), using the *hclust* function from the R
305 package STATS. Pairwise genetic differentiation (F_{ST}) among populations was estimated from
306 read count data using the R package POOLFSTAT (Hivert, Leblois, Petit, Gautier, & Vitalis,
307 2018), and pairwise geographic distances were estimated from latitude and longitude of
308 population centroids using the R package GEOSPHERE (Hijmans, Williams, Vennes, &
309 Hijmans, 2017). We tested for patterns of isolation by distance (IBD) using transformed
310 geographical (ln) and genetic distances ($F_{ST}/(1 - F_{ST})$; Rousset, 1997) and 999 permutations
311 in a Mantel test with the R package VEGAN (Oksanen, Blanchet, Kindt, Legendre, & O'Hara,
312 2011). We assessed exome-wide genetic diversity within each population by calculating the
313 proportion of polymorphic loci (PPL) and expected heterozygosity (H_e) based on population
314 allele frequencies (Fischer et al., 2017). To identify populations with substantially high and
315 low genetic diversity, we checked whether each diversity metric was normally distributed
316 using the *shapiro.test* function and identified which populations were beyond the confidence
317 interval (CI, at 97.5 % level) from a Student's *t* distribution using the *qt* function in R.
318

Compilation of different SNP sets and testing relationships

The main aim of our study was to correlate GP and HS with different measures of genetic diversity at all, neutral, and adaptive loci, respectively. To do so, we compiled ten different SNP sets (Table 1). The first set (SNP_all) included all available SNPs. Next, we identified putatively adaptive SNPs based on four different criteria (Tajima's D, π, F_{ST} outliers, SNPs associated to environmental factors) to create four different adaptive SNP sets, respectively (SNP_adaptive_D, SNP_adaptive_pi, SNP_adaptive_XTX, and SNP_adaptive_LFMM). By means of the four adaptive SNP sets, we created four neutral SNP sets (SNP_neutral_D, SNP_neutral_pi, SNP_neutral_XTX, and SNP_neutral_LFMM), which were complementary to the four adaptive SNP sets. Finally, we created a neutral SNP set (SNP_neutral_overall), which consisted of SNPs that were not included in any of the four adaptive SNP sets.

For each SNP set, we calculated four population-specific measures of genetic diversity: PPL, H_e (calculation see above), π and Θ_W (calculation see below). To identify significant relationships between GP or HS and these genetic diversity measures, we compared three nested models using analysis of variance (ANOVA) with the R package ANOVA (R Core Team, 2019); a null model (intercept only), a model adding GP or HS as linear term, and a model that additionally added GP or HS as quadratic term. The quadratic term was added to investigate non-linear response curves. We also calculated Pearson's correlation coefficients r between GP or HS and the four genetic diversity indices using the R package STATS to indicate the direction of the relationship. Furthermore, we tested correlations between diversity indices for both the full (SNP_all) and overall neutral (SNP_neutral_overall) SNP datasets using the same procedure.

342 *Gene diversity measures*

343 We calculated nucleotide diversity (π; Nei & Li, 1979) and Watterson's Θ_W (Watterson,
344 1975) to estimate Tajima's *D* (Tajima, 1989) for every contig in each population. These
345 calculations were done to identify contigs (genes) under positive selection (using π and *D*)
346 and to estimate exome-wide genetic diversity for each of the ten SNP sets (π and Θ_W). We re-
347 implemented the Python workflow used for Pool-Seq data in Sailer et al. (2018) in R and
348 performed calculations based on read count data using both variant and invariant sites at the
349 contig level. To identify genes under positive selection, we used the following procedure. For
350 π, we defined a gene as being under positive selection if the standard deviation (SD) for this
351 gene across all populations was above the 95% quantile. By doing so, we wanted to detect
352 those genes that showed the highest variation in π across populations, indicating that some,
353 but not all populations showed low gene-specific genetic diversity for some of the populations
354 compared to others, i.e. exhibited strong signatures of selection. For Tajima's *D*, we defined
355 those genes as being under positive selection which exhibited a *D* below the 5% quantile in at
356 least one population-specific distribution. We also checked for genes that were repeatedly
357 found as being under selection based on *D* across populations, which informed about the
358 proportion of common adaptive signals. For those genes identified with SD of π, we tested the
359 distribution of π values between eastern and western lineages using a Wilcoxon test in R to
360 ensure no demographic bias in gene selection. To calculate exome-wide genetic diversity for
361 the ten different SNP sets, we averaged the diversity measures of all respective genes for each
362 population. Low exome-wide values for π and Θ_W are considered indicative of low overall
363 diversity.
364

365 *F_{ST} outlier test*

366 We performed an F_{ST} outlier test to identify overly-differentiated loci using a Bayesian
367 hierarchical model implemented in BAYPASS (Gautier, 2015), which evaluates the degree of
368 differentiation of each SNP based on the X^TX genetic differentiation statistic (Günther &
369 Coop, 2013). This method accounts for pool size and read depth in Pool-Seq data and controls
370 for population genetic structure using the scaled covariance matrix of population allele
371 frequencies (Ω). We analysed the read count data of the full SNP set under the core model
372 and set the parameter d0yij to 5 (a fifth of the minimum pool size, as recommended by
373 Gautier, 2015). Then, we used a pseudo-observed data (POD) analysis to calibrate the X^TX
374 differentiation estimates and considered putatively adaptive SNPs with X^TX >99% POD
375 significant threshold. We performed 10 independent runs (with different initial seeds) and
376 computed the median of the differentiation estimates. We inspected the congruence of the
377 posterior estimates of Ω with pairwise Förstner and Moonen distances (FMD; Förstner &
378 Moonen, 2003) between the estimates of independent runs and the median. Finally, we
379 retained F_{ST} outlier loci that were identified as such in all runs and excluded those
380 inconsistently supported among runs.
381
382 *Environmental association analyses (EAAs)*
383 In EAAs (Rellstab et al., 2015), we tested for linear correlations between allele frequencies
384 and environmental variables using latent factor mixed models (LFMM; Frichot, Schoville,
385 Bouchard, & François, 2013). This approach has shown to be robust for detecting candidate
386 loci putatively under selection (De Villemereuil, Frichot, Bazin, François, & Gaggiotti, 2014;
387 Lotterhos & Whitlock, 2015) by accounting for population genetic structure with latent
388 factors in combination with test statistics to stringently control for false discoveries (François,

Martins, Caye, & Schoville, 2016). We analysed allele frequencies of the full SNP set with the function *lfmm_ridge* from $K = 2$ to $K = 8$ for each standardised (average = 0, SD = 1) environmental variable, using LFMM 2.0 implemented in the R package LFMM (Caye, Jumentier, Lepeule, & François, 2019). Genomic inflation factors (λ) were assessed with the function *lfmm_ridge* for each K value. Then, the z scores were calculated with the function *lfmm_test*, and p values were adjusted based on λ and the χ^2 distribution (Caye et al., 2019). To control for false positives, we applied the Benjamini-Hochberg algorithm with a false discovery rate (FDR) of 0.01 (Benjamini & Hochberg, 1995). We also extracted the β coefficient (regression slope) of each association and calculated the average absolute β per environmental variable for all and for the significant associations to estimate average effect sizes. We finally assembled a list of candidate loci for each environmental variable based on the optimal K value. A gene was considered adaptive if at least one of its SNPs was associated to at least one of the 34 environmental factors. Note that we extracted topographic variables for each georeferenced individual tree and averaged variables from the 20 individuals of each population to capture spatial heterogeneity and to match genetic data produced at the population level.

Results

Geographic peripherality and habitat suitability

Geographic peripherality (GP) varied from 1 km (population CH-150) to 48.1 km (CH-035) among populations (Table S5), with an average of 25.7 km (SD ± 11.7 km). Habitat suitability (HS) largely differed among populations, varying from 0.243 (CH-035) to 0.941 (CH-113) for the weighted average (Table S5). Standard deviation between models was highest for CH-052 (0.235) and lowest for CH-113 (0.065; Table S5). HS prediction across

413 the species' range was consistent among models with a moderate SD distributed across space
414 (0-0.5; Figures 3a,b, S2). Cross-validation per model resulted in high average TSS (0.882-
415 0.904; Table S6). Yearly mean temperature (Bio1) was clearly the most important variable in
416 SDMs (50.2%; Table S7), and four other variables showed an importance of at least 5%:
417 precipitation of driest quarter (Bio17, 15.9%), temperature seasonality (Bio4, 7.6%),
418 precipitation of wettest month (Bio13, 5.7%), and downslope distance gradient (t06_ddg,
419 5.4%). Overall, climatic variables were far more important in describing HS compared to
420 topographic variables (on average 11.3% compared to 1.4%; Table S7). GP and HS were
421 moderately and negatively correlated (r = -0.430, p < 0.036; Figure 3c), which allowed us to
422 independently assess correlations of GP and HS with genetic diversity.
423
424 *Exome capture sequencing and SNP detection*
425 Exome capture sequencing yielded 2.891 billion read pairs from the 24 population pools
426 (Table S4). After adapter and quality trimming, 94.0% of these reads were retained. From the
427 24 libraries, 64.5% (range: 59.0–72.2%; Table S4) of the raw read pairs mapped back to the
428 targeted transcripts. We obtained 33,125 SNPs and 3,868,577 invariant sites located in 4,870
429 single-copy genes/contigs. After missing data and MAF filtering, we retained 17,061 SNPs
430 and 3,719,732 invariant sites in 4,677 genes/contigs (Table 1), with an average of 3.6 SNPs
431 and 798.3 invariant sites per contig (range of contig size = 187–3,092 bp, median size = 723
432 bp).
433
434 *Population genetic structure and diversity*
435 The overall population genetic structure using the full SNP set (SNP_all) was consistent
436 between the hierarchical clustering tree based on Ω and the principal component analysis

437 (PCA) along the first three axes (Figures 4a, S3). In the PCA, the two main
438 phylogeographical lineages (East and West) were separated along the first principal
439 component (PC1, explaining 12.3% of the variance), and substructure in eastern and central
440 populations (i.e. contact zone of the two lineages) was revealed by PC2 (6.7%) and PC3
441 (6.0%), respectively. In total, the first four PCs summarised 30.0% of the allele frequency
442 variation among populations. Pairwise genetic differentiation between populations was low
443 overall (global F_{ST} = 0.058), with a range of pairwise F_{ST} values of 0.022 to 0.117, and
444 highest values for the EN-HJ population (Table S8). Isolation by distance was relatively high
445 and significant (Mantel r = 0.450, p < 0.001; Figure S4).
446 Overall genetic diversity (SNP_all) was similar among populations, with the
447 proportion of polymorphic loci (PPL) ranging from 0.811 to 0.912 (average 0.857) and
448 expected heterozygosity (H_e) ranging from 0.208 to 0.235 (average 0.224; Table 2). All four
449 genetic diversity estimates (PPL, H_e, π, and Θ_W) were consistently below the CI in seven
450 (CH-005, CH-008, CH-019, CH-045, CH-052, EN-HJ, and WC-HJ) and above the CI in four
451 (CH-015, CH-150, EC-HJ, and ES-HJ) populations (Table 2; Figure 4b). All diversity
452 estimates were significantly and highly correlated in both the full (SNP_all) and the overall
453 neutral SNP sets (SNP_neutral_overall; Table S9). Average values for PPL, H_e, and π were
454 slightly higher in the overall neutral SNP set (SNP_neutral_overall) compared to the full SNP
455 set (SNP_all), but lower for Θ_W (Table 2). For H_e and π, we found significantly higher genetic
456 diversity estimates in eastern compared to western populations (Figure 4c). At the exome-
457 wide level (SNP_all), Tajima's D estimates varied between 0.356 (WS-HJ) and 0.455 (CH-
458 015), with an average of 0.413 across populations (Table S10), which is compatible with a
459 past decrease in population size across the whole Swiss range of the species.
460

Gene diversity-based signature of selection

At the single-gene level and based on π and Tajima's *D*, respectively, we found 234 and 1,557 contigs as being under selection in at least one population. Of the latter, 476 contigs (30.6%) were identified as being under selection only in a single population (Figure S5), indicating that a large proportion of adaptive signals were population-specific. In turn, only 62 contigs (4.0%) were found as being under selection in at least half (12) of the sampled populations (Figure S5). In total, 169 (3.6%) of the 4,677 contigs showed a strong signature of selection in both π and Tajima's *D*. The π values of the two phylogeographical lineages were not significantly different (Figure S6).

F_{ST} outlier test

Analysis of the full dataset (SNP_all) under the BAYPASS core model (X^TX) revealed that 205 SNPs from 154 contigs were overly differentiated among populations and putatively exhibited signals of adaptive divergence. Pairwise FMDs between independent runs and their median were lower than 0.072 (SD ± 0.004), and topologies of the hierarchical clustering trees (HCT) generated from the dissimilarity matrix Ω were unchanged among runs. For the POD, pairwise FMDs between independent runs and the median were low (0.847 ± 0.034 SD), and topologies of the HCT showed slight differences. Pairwise FMDs between the median of the original posterior estimates of Ω and the one calculated from the POD was higher (5.670) and stable across the different runs.

Environmental association analyses

In LFMM, the genomic inflation factor (λ) differed slightly among *K* values and was on average lowest for $K = 3$ (Table S11). Based on this optimal *K* value, we found a total of 625

485 significant associations of a SNP with one of the 34 environmental variables (Table S12).
486 This number of associations included 346 different SNPs that represent 2.0% of the exome-
487 wide SNP set. From these SNPs, 189 (54.6%) were associated with a single environmental
488 variable and 157 (45.4%) with at least two variables. The number of significant associations
489 largely differed among environmental variables, from 0 for several variables to 117
490 associations for precipitation of the warmest quarter (Bio18; Table S12). Apart from Bio18,
491 temperature seasonality (Bio4) and temperature annual range (Bio7) showed the highest
492 numbers of significant associations (88 and 80, respectively). Note that in some cases, despite
493 a high Pearson's correlation (r >|0.7|) between environmental variables, contrasting numbers
494 of significant associations were found (e.g. Bio4 and Bio11; Figure S7). The ranking of these
495 variables, either based on the number of significant associations or the averaged absolute β
496 coefficients, was not significantly correlated with the one representing the variable
497 importance in SDM (Table S12; Figure S8). Overall, we found more average climate- (25.7)
498 than topography-related (8.9) associations per variable (Table S12).
499
500 *Relationship between geographic peripherality, habitat suitability, and genetic diversity*
501 For the full (SNP_all) and all five neutral SNP sets, GP was significantly and negatively
502 correlated with all genetic diversity indices (Table 3), i.e. peripheral populations tended to
503 have lower genetic diversity than central populations (the example of H_e of
504 SNP_neutral_overall is given in Figure 5a). None of the diversity indices was correlated to
505 HS in the full and all neutral datasets (example given in Figure 5b).
506 Most of the correlations of GP or HS with genetic diversity at adaptive loci were not
507 significant (the example of H_e of SNP_adaptive_D for GP is given in Figure 5c). However,
508 there were three exceptions. PPL and Θ_W were negatively correlated with GP in the adaptive

SNP set based on π (SNP_adaptive_pi), and H_e in the adaptive SNP set based on low population-specific *D* values (SNP_adaptive_D) was negatively correlated with HS (Figure 5d). Hence, populations in more suitable habitats tended to have lower genetic diversity at adaptive loci than populations towards the margin of the niche.

Discussion

A better understanding of the potential key drivers of genetic diversity at neutral and adaptive loci is essential for the assessment of a species' adaptive capacity (Flanagan, Forester, Latch, Aitken, & Hoban, 2018). Our study provides one of the first empirical comparisons of genetic diversity at both neutral and putatively adaptive loci in relation to geographic position within the species' native range (i.e. geographic peripherality, GP) and environmental conditions (i.e. habitat suitability, HS). The fact that GP and HS were only moderately correlated allowed us to disentangle these two drivers of genetic diversity (Figure 3c). We combined species distribution models with exome-wide polymorphism data (17,061 SNPs from 4,677 contigs/genes of the estimated 30,000–50,000 genes that can be identified in conifers; Neale & Wheeler, 2019), and found that neutral genetic diversity was negatively correlated with the distance to the range centre (i.e. GP); populations living at the periphery of the distribution had lower neutral or overall genetic diversity than populations from the central area (Table 3). In contrast, neutral genetic diversity was not correlated with HS. Moreover, estimates of genetic diversity at adaptive loci were also not correlated with GP or HS in most cases. However, in the adaptive SNP set based on Tajima's *D*, heterozygosity (H_e) was negatively correlated with HS, meaning that populations situated in less suitable habitats had a higher genetic diversity at adaptive loci than populations in more suitable habitats (Table 3; Figure 5d). This finding agrees with our expectation under the assumption that migration is weaker

than selection (solid line in Figure 1d). Based on these insights, we highlight the importance of distinguishing neutral from adaptive genetic variation.

Geographic peripherality, habitat suitability, and genetic diversity at neutral loci

Our results based on neutral or overall genetic SNP sets agree with the hypothesised pattern in respect to GP (Figure 1a, 5a). Neutral genetic diversity is reduced at the range limit as compared to the central areas of occurrences, which is consistent with the known population census sizes of *P. cembra* (Fischer & Traub, 2019). Surprisingly, populations at the contact zone between the two main phylogeographic lineages of *P. cembra* in the Swiss range did not exhibit higher values of heterozygosity than the rest of populations, which could be expected as a result of admixture (Figure 4). The higher genetic diversity found in eastern compared to the western populations (Figure 4c) is consistent with the presumed main re-colonisation route of *P. cembra* from the eastern fringe of the Alps into its current Alpine range after the last glacial maximum (Gugerli et al., 2009). In turn, a presumed second immigration route advanced eastward from a likely smaller refugial area in or near the western Alps (Tóth, Tremblay, Housset, Bergeron, & Carcaillet, 2019), possibly complemented by a third lineage entering the central Alps from the South (Vescovi et al., 2007). The geographic position of populations *per se* unlikely affects neutral genetic diversity, but rather acts via effective population size N_e (through strength of genetic drift, hence reduction in heterozygosity) and population connectivity (through gene flow). The highest measures of neutral genetic diversity found in our study suggest that, since the onset of post-glacial re-colonisation of the central Alps by *P. cembra* (i.e. about 10,000 years ago; Vescovi et al., 2007), the Alpine meta-population has carried over a large amount of standing genetic variation from eastern, southern and western refugia to the current central populations.

557 Assuming a generation time of about 50 years for *P. cembra* (Zoller, 1991), at least 200
558 generations have passed since the central Alpine populations became established, which
559 seems to be a rather limited turn-over to accumulate mutations and homogenise allele
560 frequencies across populations (Austerlitz, Mariette, Machon, Gouyon, & Godelle, 2000).
561 Interestingly, despite significant isolation by distance among populations, the low pairwise
562 genetic differentiation (global F_{ST} = 0.058; Table S8) suggests that either historical gene flow
563 or to a lesser extent standing genetic variation is relevant over the entire range of the study
564 species within the Swiss Alps.
565 In contrast to GP, HS was not correlated with genetic diversity at neutral loci. This
566 finding does not confirm the hypothesis presented in Figure 1b, which assumes that
567 geographic and environmental features are highly correlated. While this was not the case in
568 our study, one could argue that the environment affects neutral genetic diversity indirectly
569 through effective population size N_e. Our observations are rather consistent with patterns of
570 genetic diversity recently reported for *P. cembra* in a regional study from the southwestern
571 Alps, where marginal populations harboured similar neutral genetic diversity as core
572 populations (Tóth et al., 2019).
573
574 *Geographic peripherality, habitat suitability, and genetic diversity at adaptive loci*
575 Contrarily to the neutral and overall SNP sets, there was no significant relationship
576 between GP and diversity indices in the adaptive SNP sets. Generally, our results confirm the
577 pattern hypothesised (Figure 1c), suggesting that geographic features have no effect on
578 genetic diversity at adaptive loci, also not indirectly through the effective population size N_e.
579 However, there were two cases of genetic diversity estimates (PPL and Θ_W) showing a
580 negative correlation between genetic diversity and GP in the adaptive SNP set based on π

581 (SNP_adapative_pi; Table 3). This SNP set shows a similar pattern as neutral SNP sets, but to
582 a lesser degree. This finding could imply that our method to identify adaptive SNPs based on
583 relative measures (maximum SD) of π led to a SNP set that still contained a considerable
584 proportion of neutral SNPs or also SNPs under balancing selection in single populations
585 (Figure S6a). However, using a lower quantile threshold of absolute values of π would not
586 improve the selection of genes, because it would target mostly genes with no variation (π =
587 0), which likely are the result of purifying, and not positive selection.
588 Most correlations between HS and genetic diversity at adaptive loci were also not
589 significant. However, the significant negative relationship between HS and H_e for the
590 adaptive SNP set based on Tajima's *D* indicates a continuum of selection responses along
591 habitat conditions. Focusing on this significant relationship, we see that the strongest signals
592 of positive selection are detected in populations living in highly suitable habitats, where
593 among-population gene flow seemed not to fully counteract selection processes, indicating
594 that m < s as depicted by the solid line in Figure 1d. Populations at the core of the niche may
595 have gone through a strong selection process, hence currently harbour the best suited allele
596 composition, and diversity at adaptive loci is therefore reduced. Low genetic diversity at
597 adaptive loci is, however, only the case for single populations; overall genetic diversity at
598 adaptive loci across all populations might still be high as supported by the fact that many
599 adaptive signals (detected with Tajima's *D*) are population-specific (Figure S5). In other
600 words, populations have presumably developed independent molecular solutions for adapting
601 to similar environmental conditions (Rellstab et al., 2017). Note that genetic drift and allele
602 surfing can potentially mimic such a reduction in genetic diversity at some loci (e.g. Excoffier
603 & Ray, 2008), but presumably not at the level of the representative fraction of adaptive gene
604 space. Conversely, populations at the niche margin might still be in the process of locally

605 adapting, relaxing selection, or gene flow from differently adapted populations, which leads
606 to immigrating of mal-adapted alleles, hence contain higher genetic diversity at these adaptive
607 loci.

608 Most importantly, our results show that it is important to distinguish between genetic
609 diversity at neutral and adaptive loci when investigating the geographic and environmental
610 drivers of genetic diversity. Our investigation also indicates that in sampling designs with
611 thousands of SNPs, using the whole SNP set (e.g. Lee-Yaw et al., 2018) may lead to similar
612 results as using neutral loci only (Table 2), even in an exome capture sequencing approach
613 that mainly targets coding regions (i.e. possible targets of natural selection). Unfortunately,
614 most studies focusing on population genetic diversity, have ignored the distinction between
615 neutral and adaptive loci so far (but see Aguirre-Liguori et al., 2017). This is partly due to the
616 fact that it was technically difficult to discriminate between neutral and adaptive genetic
617 diversity, because next-generation sequencing (NGS) techniques or genomic resources of
618 non-model species were not yet available. However, costs for NGS are steadily decreasing
619 (https://www.genome.gov/sequencingcostsdata), reaching reasonable amounts for reduced-
620 representation sequencing approaches like exome capture (Yeaman et al., 2016) or RAD-Seq
621 (Andrews, Good, Miller, Luikart, & Hohenlohe, 2016), even in species with large genomes
622 like *P. cembra*. Moreover, access to high-performance computer clusters is now available and
623 user-friendly bioinformatic software is being developed (Danecek et al., 2011; Puritz,
624 Hollenbeck, & Gold, 2014; Van der Auwera et al., 2013). One major challenge remains, i.e.,
625 the identification of genes involved in adaptive processes to distinguish between neutral and
626 adaptive genetic variation, a task greatly assisted by the increasing number of available,
627 annotated reference genomes (e.g. Lewin et al., 2018; Twyford, 2018). In the present study,
628 we utilised a suite of approaches to identify putatively adaptive loci, but the task remains

imperfect as a consequence of, e.g., false positives and negatives, arbitrary thresholds, missing functional annotation, population-specific signatures of selection, and polygenic processes with many small-effect loci. From the results of our empirical study, we recommend to disentangle neutral and adaptive genetic variation as far as possible for a better understanding of a species' demographic and adaptation history. Moreover, we recall the importance of using complementary approaches for detecting signatures of selection—i.e. including population-specific analyses (π and Tajima's D in our study), F_{ST} outlier tests or EAA—because a single method might fail to well describe the diverse signatures of adaptation (Hohenlohe, Phillips, & Cresko, 2010).

Environmental factors in species distribution models and environmental association analyses

Although yearly mean temperature (Bio1) was the most important variable for predicting *P. cembra*'s habitat suitability (Table S7), this variable did not show a large number of significant associations, nor a high effect size in EAA (Table S12). Likewise, variable importance in the SDMs did not correlate with EAA-based importance parameters (Figure S8). This suggests that variable selection in adaptation studies should not be done using *a priori* knowledge solely based on the power of a variable to predict a species' realised ecological niche. In other words, it is challenging to obtain relevant clues of selective forces at the local scale when habitat characterisation depends on ecological data from the entire species' range (but see, e.g., Borrell, Zohren, Nichols, & Buggs, 2020), especially if the study design consists of a partial sampling at its leading or rear edges (Hampe & Petit, 2005). One reason that might explain this mismatch is the temporal lag involved in the two processes; species presence can reflect rather recent events, while selection signatures are related to an evolutionary time scale, whose dimension depends, among others, on the species' generation

time. Moreover, a species that is highly adapted to a certain niche (e.g. high-altitude habitats in the case of *P. cembra*) may experience a limited range in certain environmental factors. The detection of selection signatures in the genome, however, is often increased in environmentally heterogenous study systems (Lotterhos & Whitlock, 2015), potentially leading to the observed mismatch in variable importance. Another reason could be that yearly mean temperature might well define *P. cembra*'s realised niche limits at the cold or warm ends of the temperature gradient, while local adaptation within these general limits acts along other gradients (such as temperature seasonality or summer precipitation).

Conclusions

Genetic diversity is a key feature in ecology and evolution, because it is (i) an important part of biodiversity, and (ii) considering that adaptive and neutral genetic diversity involve distinct biological processes, it is a suitable proxy of population resilience under environmental change. With access to large genomic datasets from geo-referenced populations and individuals, in combination with new spatial and statistical tools, it is now possible to distinguish genetic diversity at adaptive and neutral loci, despite the confounding signals of adaptation processes and population demographic history. This is of special importance for conservation prospects, in which forest and conservation managers or other stakeholders need reliable estimates of population genetic diversity at adaptive loci together with an assessment of associated uncertainties for drawing recommendations in the context of environmental change.

Acknowledgements

We thank L. Ammann, B. Bauert, M. Coleman, D. Galvan, R. Graf, L. Hellmann, S. Klesse, B. Lendvay, J. Marchand, R. Meier, J. Müller, D. Nievergelt, A. Rogivue, M. Schmid and R. Winiger for help during field work, forest owners and services for sampling permissions, the staff at Fasteris (Geneva) and the Functional Genomics Center Zurich (FGCZ) for Illumina sequencing, and C. Sailer for support in translating his Python workflow into R. Constructive comments from two anonymous reviewers are also acknowledged. This study was funded by the Swiss National Science Foundation through grant 31003A_152664 (awarded to F.G.).

References

Aguirre-Liguori, J. A., Tenaillon, M. I., Vázquez-Lobo, A., Gaut, B. S., Jaramillo-Correa, J. P., Montes-Hernandez, S., … Eguiarte, L. E. (2017). Connecting genomic patterns of local adaptation and niche suitability in teosintes. *Molecular Ecology*, *26*(16), 4226–4240. doi:10.1111/mec.14203

Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T., & Curtis-McLane, S. (2008). Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, *1*(1), 95–111. doi:10.1002/2016SW001410

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, *43*(6), 1223–1232. doi:10.1111/j.1365-2664.2006.01214.x

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*(2), 1–30. doi:http://dx.doi.org/10.1101/028837

Araújo, M. B., Anderson, R. P., Barbosa, A. M., Beale, C. M., Dormann, C. F., Early, R., … Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, *5*(1), 1–12. doi:10.1126/sciadv.aat4858

Austerlitz, F., Mariette, S., Machon, N., Gouyon, P. H., & Godelle, B. (2000). Effects of colonization processes on genetic diversity: Differences between annual plants and tree species. *Genetics*, *154*(3), 1309–1321.

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology and Evolution*. doi:10.1016/j.tree.2007.09.008

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Royal Statistical Society*, *57*(1), 289–300.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. doi:10.1093/bioinformatics/btu170

Borrell, J. S., Zohren, J., Nichols, R. A., & Buggs, R. J. (2020). Genomic assessment of local adaptation in dwarf birch to inform assisted gene flow. *Evolutionary Applications*, *13*(1),

161–175. doi:10.1111/eva.12883
Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. doi:10.3390/rs10060911
Brown, J. H. (1984). On the relationship between abundance and distribution of species. *The American Naturalist*, *124*(2), 255–279. doi:10.1093/ehr/cepl85
Caudullo, G., Welk, E., & San-Miguel-Ayanz, J. (2017). Chorological maps for the main European woody species. *Data in Brief*, *12*, 662–666. doi:10.1016/j.dib.2017.05.007
Caye, K., Jumentier, B., Lepeule, J., & François, O. (2019). LFMM 2: Fast and accurate inference of gene-environment associations in genome-wide studies. *Molecular Biology and Evolution*, *36*(4), 852–860. doi:10.1093/molbev/msz008
Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., … Böhner, J. (2015). System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, *8*(7), 1991–2007. doi:10.5194/gmd-8-1991-2015
Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. doi:10.1093/bioinformatics/btr330
De Villemereuil, P., Frichot, É., Bazin, É., François, O., & Gaggiotti, O. E. (2014). Genome scan methods against more complex models: When and how much should we trust them? *Molecular Ecology*, *23*(8), 2006–2019. doi:10.1111/mec.12705
Eckert, C. G., Samis, K. E., & Lougheed, S. C. (2008). Genetic variation across species' geographical ranges: The central-marginal hypothesis and beyond. *Molecular Ecology*, *17*(5), 1170–1188. doi:10.1111/j.1365-294X.2007.03659.x
Excoffier, L., & Ray, N. (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology and Evolution*, *23*(7), 347–351. doi:10.1016/j.tree.2008.04.004
Exposito-Alonso, M., Burbano, H. A., Bossdorf, O., Nielsen, R., & Weigel, D. (2019). Natural selection on the *Arabidopsis thaliana* genome in present and future climates. *Nature*, *573*(7772), 126–129. doi:10.1038/s41586-019-1520-9
Felsenstein, J. (1976). The theoretical population genetics of variable selection and migration. *Annual Review of Genetics*, *10*(1), 253–280. doi:10.1146/annurev.ge.10.120176.001345
Fischer, C., & Traub, B. (2019). *Swiss National Forest Inventory – Methods and Models of the Fourth Assessment*. (C. Fischer & B. Traub, Eds.) (Fourth). WSL, Birmensdorf: Springer.
Fischer, M. C., Foll, M., Heckel, G., & Excoffier, L. (2014). Continental-scale footprint of balancing and positive selection in a small rodent (Microtus arvalis). *PLoS One*, *9*(11). doi:10.1371/journal.pone.0112332
Fischer, M. C., Rellstab, C., Leuzinger, M., Roumet, M., Gugerli, F., Shimizu, K. K., … Widmer, A. (2017). Estimating genomic diversity and population differentiation - an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. *BMC Genomics*, *18*(1), 1–15. doi:10.1186/s12864-016-3459-7
Flanagan, S. P., Forester, B. R., Latch, E. K., Aitken, S. N., & Hoban, S. (2018). Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. *Evolutionary Applications*, *11*(7), 1035–1052. doi:10.1111/eva.12569
Förstner, W., & Moonen, B. (2003). A metric for covariance matrices. In E. W. Grafarend, F. W. Krumm, & V. S. Schwarze (Eds.), *Geodesy-The Challenge of the 3rd Millennium* (pp. 299–309). Berlin, Germany: Springer. doi:10.1017/CBO9781107415324.004
François, O., Martins, H., Caye, K., & Schoville, S. D. (2016). Controlling false discoveries

761 in genome scans for selection. *Molecular Ecology*, *25*(2), 454–469.
762 doi:10.1111/mec.13513
763 Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations
764 between loci and environmental gradients using latent factor mixed models. *Molecular*
765 *Biology and Evolution*, *30*(7), 1687–1699. doi:10.1093/molbev/mst063
766 Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with
767 population-specific covariates. *Genetics*, *201*(4), 1555–1579.
768 doi:10.1534/genetics.115.181453
769 Gernandt, D. S., Geada López, G., Ortiz García, S., & Liston, A. (2005). Phylogeny and
770 classification of *Pinus*. *Taxon*, *54*(1), 29–42. doi:10.2307/25065300
771 Gravel, S. (2016). When is selection effective? *Genetics*, *203*(1), 451–462.
772 doi:10.1534/genetics.115.184630
773 Gugerli, F., Rüegg, M., & Vendramin, G. G. (2009). Gradual decline in genetic diversity in
774 Swiss stone pine populations (*Pinus cembra*) across Switzerland suggests postglacial re-
775 colonization into the Alps from a common eastern glacial refugium. *Botanica Helvetica*,
776 *119*(1), 13–22. doi:10.1007/s00035-009-0052-6
777 Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology.
778 *Ecological Modelling*, *135*(2–3), 147–186. doi:10.1016/S0304-3800(00)00354-9
779 Günther, T., & Coop, G. (2013). Robust identification of local adaptation from allele
780 frequencies. *Genetics*, *195*(1), 205–220. doi:10.1534/genetics.113.152462
781 Hampe, A., & Petit, R. J. (2005). Conserving biodiversity under climate change: The rear
782 edge matters. *Ecology Letters*, *8*(5), 461–467. doi:10.1111/j.1461-0248.2005.00739.x
783 Hastie, T., & Tibshirani. (2017). Exploring the nature of covariate effects in the proportional
784 hazards model. *Biometrics*, *46*(4), 1005–1016.
785 Hengelveld, R., & Haeck, J. (1982). The distribution of abundance. I. Measurements. *Journal*
786 *of Biogeography*, *9*(4), 303–316.
787 Hijmans, R. J., Phillips, S., Leathwick, J., Elith, J., & Hijmans, M. R. J. (2017). Package
788 dismo. *Circles*, *9*(1), 1–68.
789 Hijmans, R. J., Williams, E., Vennes, C., & Hijmans, M. R. J. (2017). Introduction to the
790 geosphere R package. R package 1.5.10. Retrieved from
791 www.rspatial.org/sphere/sphere.pdf
792 Hivert, V., Leblois, R., Petit, E. J., Gautier, M., & Vitalis, R. (2018). Measuring genetic
793 differentiation from pool-seq data. *Genetics*, *210*(1), 315–330.
794 doi:10.1534/genetics.118.300900
795 Hohenlohe, P. A., Phillips, P. C., & Cresko, W. A. (2010). Using population genomics to
796 detect selection in natural populations: Key concepts and methodological considerations.
797 *International Journal of Plant Sciences*, *171*(9), 1059–1071. doi:10.1086/656306
798 Höhn, M., Gugerli, F., Abran, P., Bisztray, G., Buonamici, A., Cseke, K., … Vendramin, G.
799 G. (2009). Variation in the chloroplast DNA of Swiss stone pine (*Pinus cembra* L.)
800 reflects contrasting post-glacial history of populations from the Carpathians and the
801 Alps. *Journal of Biogeography*, *36*(9), 1798–1806. doi:10.1111/j.1365-
802 2699.2009.02122.x
803 Houle, D. (1992). Comparing evolvability and variability. *Genetics*, *130*(1), 195–204.
804 doi:citeulike-article-id:10041224
805 Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-
806 efficient alignment of short DNA sequences to the human genome. *Genome Biology*,
807 *10*(3). doi:10.1186/gb-2009-10-3-r25
808 Lee-Yaw, J. A., Fracassetti, M., & Willi, Y. (2018). Environmental marginality and

809 geographic range limits: a case study with *Arabidopsis lyrata* ssp. *lyrata*. *Ecography*,
810 *41*(4), 622–634. doi:10.1111/ecog.02869
811 Leempoel, K., Parisod, C., Geiser, C., Daprà, L., Vittoz, P., & Joost, S. (2015). Very high-
812 resolution digital elevation models: Are multi-scale derived variables ecologically
813 relevant? *Methods in Ecology and Evolution*, *6*(12), 1373–1383. doi:10.1111/2041-
814 210X.12427
815 Lei, Y. K., Wang, W., Liu, Y. P., He, D., & Li, Y. (2015). Adaptive genetic variation in the
816 smoke tree (*Cotinus coggygria* Scop.) is driven by precipitation. *Biochemical*
817 *Systematics and Ecology*, *59*, 63–69. doi:10.1016/j.bse.2015.01.009
818 Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., …
819 Zhang, G. (2018). Earth BioGenome project: Sequencing life for the future of life.
820 *Proceedings of the National Academy of Sciences of the United States of America*,
821 *115*(17), 4325–4333. doi:10.1073/pnas.1720115115
822 Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, *2*(3),
823 18–22.
824 Lira-Noriega, A., & Manthey, J. D. (2014). Relationship of genetic diversity and niche
825 centrality: A survey and analysis. *Evolution*, *68*(4), 1082–1093. doi:10.1111/evo.12343
826 Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect
827 local adaptation depends on sampling design and statistical method. *Molecular Ecology*,
828 *24*(5), 1031–1046. doi:10.1111/mec.13100
829 Macdonald, S. L., Llewelyn, J., Moritz, C., & Phillips, B. L. (2017). Peripheral isolates as
830 sources of adaptive diversity under climate change. *Frontiers in Ecology and Evolution*,
831 *5*(August), 1–10. doi:10.3389/fevo.2017.00088
832 Martínez-Meyer, E., Díaz-Porras, D., Peterson, A. T., & Yáñez-Arenas, C. (2013). Ecological
833 niche structure and rangewide abundance patterns of species. *Biology Letters*, *9*(1),
834 20120637. doi:10.1098/rsbl.2012.0637
835 McKenna, A., Hanna, M., Banks, E., Sivachenko, A. Y., Cibulskis, K., Kernytsky, A. M., …
836 DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for
837 analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303.
838 doi:10.1101/gr.107524.110.20
839 McKinney, G. J., Waples, R. K., Seeb, L. W., & Seeb, J. E. (2017). Paralogs are revealed by
840 proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing
841 data from natural populations. *Molecular Ecology Resources*, *17*(4), 656–669.
842 doi:10.1111/1755-0998.12613
843 Meier, E. S., Kienast, F., Pearman, P. B., Svenning, J. C., Thuiller, W., Araújo, M. B., …
844 Zimmermann, N. E. (2010). Biotic and abiotic variables show little redundancy in
845 explaining tree species distributions. *Ecography*, *33*(6), 1038–1048. doi:10.1111/j.1600-
846 0587.2010.06229.x
847 Motta, R., & Nola, P. (2001). Growth trends and dynamics in sub-alpine forest stands in the
848 Varaita Valley (Piedmont, Italy) and their relationships with human activities and global
849 change. *Journal of Vegetation Science*, *12*(2), 219–230. doi:10.2307/3236606
850 Neale, D. B., & Wheeler, N. C. (2019). *The conifers: Genomes, variation and evolution*. *The*
851 *Conifers: Genomes, Variation and Evolution*. Cham: Springer. doi:10.1007/978-3-319-
852 46807-5
853 Nei, M. (2013). *Mutation-Driven Evolution*. Oxford: Oxford University Press.
854 Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of
855 restriction endonucleases. *Proceedings of the National Academy of Sciences of the*
856 *United States of America*, *76*(10), 5269–5273. doi:10.1073/pnas.76.10.5269

Nei, M., Suzuki, Y., & Nozawa, M. (2010). The neutral theory of molecular evolution in the genomic era. *Annual Review of Genomics and Human Genetics*, *11*(1), 265–289. doi:10.1146/annurev-genom-082908-150129

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, *135*(3), 370–384. doi:https://doi.org/10.2307/2344614

Neuschulz, E. L., Merges, D., Bollmann, K., Gugerli, F., & Böhning-Gaese, K. (2018). Biotic interactions and seed deposition rather than abiotic factors determine recruitment at elevational range limits of an alpine tree. *Journal of Ecology*, *106*(3), 948–959. doi:10.1111/1365-2745.12818

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., & O'Hara, R. B. (2011). Vegan: Community ecology package. R package 2.5.6. Retrieved from www.cran.r-project.org/web/packages/vegan/index.html

Phillips, S. B., Aneja, V. P., Kang, D., & Arya, S. P. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(2), 231–252. doi:10.1016/j.ecolmodel.2005.03.026

Puritz, J. B., Hollenbeck, C. M., & Gold, J. R. (2014). dDocent: A RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, *2*, e431. doi:10.7717/peerj.431

R Core Team. (2020). R: A language and environment for statistical computing.

Rellstab, C., Dauphin, B., Zoller, S., Brodbeck, S., & Gugerli, F. (2019). Using transcriptome sequencing and pooled exome capture to study local adaptation in the giga-genome of *Pinus cembra*. *Molecular Ecology Resources*, *19*(2), 536–551. doi:10.1111/1755-0998.12986

Rellstab, C., Fischer, M. C., Zoller, S., Graf, R., Tedder, A., Shimizu, K. K., … Gugerli, F. (2017). Local adaptation (mostly) remains local: Reassessing environmental associations of climate-related candidate SNPs in *Arabidopsis halleri*. *Heredity*, *118*(2), 193–201. doi:10.1038/hdy.2016.82

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, *24*(17), 4348–4370. doi:10.1111/mec.13322

Rellstab, C., Zoller, S., Tedder, A., Gugerli, F., & Fischer, M. C. (2013). Validation of SNP allele frequencies determined by pooled next-generation sequencing in natural populations of a non-model plant species. *PLoS One*, *8*(11), e80422. doi:10.1371/journal.pone.0080422

Ripley, B. D. (1996). *Pattern Recognition via Neural Networks*. *Cambridge University Press, Cambridge*.

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, *145*(4), 1219–1228. doi:10.1007/BF00341816

Sagarin, R. D., & Gaines, S. D. (2002). The "abundant centre" distribution: To what extent is it a biogeographical rule? *Ecology Letters*, *5*(1), 137–147. doi:10.1046/j.1461-0248.2002.00297.x

Sailer, C., Babst-Kostecka, A., Fischer, M. C., Zoller, S., Widmer, A., Vollenweider, P., … Rellstab, C. (2018). Transmembrane transport and stress response genes play an important role in adaptation of *Arabidopsis halleri* to metalliferous soils. *Scientific Reports*, *8*(1), 1–13. doi:10.1038/s41598-018-33938-2

Salzer, K. (2011). *Wind- and bird-mediated gene flow in* Pinus cembra*: effects on spatial genetic structure and potential close-relative inbreeding*. *WSL Birmensdorf*.

Salzer, K., & Gugerli, F. (2012). Reduced fitness at early life stages in peripheral versus core

populations of Swiss stone pine (*Pinus cembra*) is not reflected by levels of inbreeding in seed families. *Alpine Botany*, *122*(2), 75–85. doi:10.1007/s00035-012-0106-z

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, *14*(11), 807–820. doi:10.1038/nrg3522

Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals-mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, *15*(11), 749–763. doi:10.1038/nrg3803

Shapiro, J. A., Huang, W., Zhang, C., Hubisz, M. J., Lu, J., Turissini, D. A., … Wu, C. I. (2007). Adaptive genic evolution in the *Drosophila* genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(7), 2271–2276. doi:10.1073/pnas.0610385104

Slatkin, M. (1973). Gene flow and selection in a cline. *Genetics*, *75*(4), 733–756.

Slatkin, M. (1985). Gene flow in natural populations. *Annual Review of Ecology and Systematics*, *16*(1985), 393–430. doi:10.1146/annurev.es.16.110185.002141

Šurinová, M., Hadincová, V., Vandvik, V., & Münzbergová, Z. (2019). Temperature and precipitation, but not geographic distance, explain genetic relatedness among populations in the perennial grass *Festuca rubra*. *Journal of Plant Ecology*, *12*(4), 730–741. doi:10.1093/jpe/rtz010

Swisstopo. (2004). *DHM25 The digital height model of Switzerland*. Swiss Federal Office of Topography, Wabern.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*, 585–595. doi:PMC1203831

Thornton, P. E., Running, S. W., & White, M. A. (1997). Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology*, *190*(3–4), 214–251. doi:10.1016/S0022-1694(96)03128-9

Tóth, E. G., Tremblay, F., Housset, J. M., Bergeron, Y., & Carcaillet, C. (2019). Geographic isolation and climatic variability contribute to genetic differentiation in fragmented populations of the long-lived subalpine conifer *Pinus cembra* L. in the western Alps. *BMC Evolutionary Biology*, *190*(19), 1–17.

Twyford, A. D. (2018). The road to 10,000 plant genomes. *Nature Plants*, *4*(6), 312–313. doi:10.1038/s41477-018-0165-2

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., … DePristo, M. A. (2013). From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 1–33. doi:10.1002/0471250953.bi1110s43

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*. Springer, New York. doi:10.2307/2685660

Vescovi, E., Ravazzi, C., Arpenti, E., Finsinger, W., Pini, R., Valsecchi, V., … Tinner, W. (2007). Interactions between climate and vegetation during the Lateglacial period as recorded by lake and mire sediment archives in Northern Italy and Southern Switzerland. *Quaternary Science Reviews*, *26*(11–12), 1650–1669. doi:10.1016/j.quascirev.2007.03.005

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, *7*(2), 256–276.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *73*(1), 3–36. doi:10.1111/j.1467-9868.2010.00749.x

Wüest, R. O., Bergamini, A., Bollmann, K., & Baltensweiler, A. (2020). LiDAR data as a proxy for light availability improve distribution modelling of woody species. *Forest Ecology and Management*, *456*(September 2019), 117644. doi:10.1016/j.foreco.2019.117644

Yeaman, S., Hodgins, K. A., Lotterhos, K. E., Suren, H., Nadeau, S., Degner, J. C., … Aitken, S. N. (2016). Convergent local adaptation to climate in distantly related conifers. *Science*, *353*(6306), 23–26. doi:10.1126/science.aaf7812

Zoller, H. (1991). *Gustav Hegi—Illustrierte Flora von Mitteleuropa*. Berlin, Germany: Blackwell.

Zonneveld, B. J. M. (2012). Conifer genome sizes of 172 species, covering 64 of 67 genera, range from 8 to 72 picogram. *Nordic Journal of Botany*, *30*(4), 490–502. doi:10.1111/j.1756-1051.2012.01516.x

Data Accessibility

Raw sequence data used in this study are accessible at NCBI under SRA accessions nos. SRR8237211−SRR8237217 (EC-HJ to WZ-HJ) and at the European Nucleotide Archive (ENA) under accession nos. ERS4525650−ERS4525666 (CH-005 to CH-150). Allele frequencies and environmental datasets together with R scripts used for analyses will be uploaded to the Dryad Digital Repository upon acceptance.

Author Contributions

F.G. acquired funding. B.D., R.O.W., F.G., R.H., and C.R. designed the conceptual approach. C.R., F.G., and S.B. carried out field work. S.B. performed laboratory work. S.Z. and B.D. performed bioinformatic analyses. B.D. and R.O.W. generated and analysed topographic and climatic data. R.O.W. carried out species distribution modelling. B.D. and M.C.F analysed genomic data. B.D. wrote the manuscript, with major contributions from R.O.W., M.C.F., R.H, F.G., and C.R. All authors read, commented and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Tables

TABLE 1 Details of the full, neutral, and adaptive SNP sets generated for the studied *Pinus cembra* populations. Thresholds and the main parameters used in analyses are summarised, and numbers of contigs and SNPs are indicated.

Set	Type	Abbreviation	Description and thresholds used in analyses	# Contigs	# SNPs
1	All	SNP_all	All SNPs	4,677	17,061
2	Neutral	SNP_neutral_D	All SNPs excluding SNP_adaptive_D	3,120	9,602
3	Neutral	SNP_neutral_pi	All SNPs excluding SNP_adaptive_pi	4,443	15,273
4	Neutral	SNP_neutral_XTX	All SNPs excluding SNP_adaptive_XTX	4,651	16,856
5	Neutral	SNP_neutral_LFMM	All SNPs excluding SNP_adaptive_LFMM	4,648	16,717
6	Neutral	SNP_neutral_overall	All SNPs without any adaptive signature (excluding SNP sets 7-10)	8,802	3,007
7	Adaptive	SNP_adaptive_D	SNPs in genes below the 0.05 quantile of *D* in at least one population	232 – 262	1,254 – 1,437
8	Adaptive	SNP_adaptive_pi	SNPs in genes above the 0.95 quantile of the standard deviation of (π) across all populations	234	1,788
9	Adaptive	SNP_adaptive_XTX	F_{ST} outlier SNPs in BAYPASS ($X^TX > 0.99$ POD)	154	205
10	Adaptive	SNP_adaptive_LFMM	SNPs significantly associated to environmental factors in LFMM (FDR < 0.01)	221	346

TABLE 2 Summary of four population genetic diversity estimates for 24 *Pinus cembra* populations for the full (SNP_all) and overall neutral (SNP_neutral_overall) SNP sets. PPL: proportion of polymorphic loci, H_e: expected heterozygosity, π: nucleotide diversity, Θ_W: Watterson's theta. Values below and above the 97.5% confidence interval (CI) of the *t* distribution are represented in italics and in bold, respectively. Populations were classified as part of the eastern or western lineage of *P. cembra* based on Figure 4a and in agreement with Gugerli et al. (2009).

			Full SNP set				**Overall neutral SNP set**			
Population	**Lineage**	**Sample**	**PPL**	H_e	π	Θ_W	**PPL**	H_e	π	Θ_W
Chandolin	Western	CH-005	*0.844*	*0.221*	*1.192E-03*	*0.657*	*0.881*	0.278	1.252E-03	0.578
Forêt du Lapé	Western	CH-008	*0.812*	*0.214*	*1.154E-03*	*0.635*	*0.869*	*0.272*	*1.221E-03*	*0.567*
Avers	Eastern	CH-011	0.864	0.225	1.217E-03	0.683	0.894	0.280	1.257E-03	0.587
Tamangur	Eastern	CH-015	**0.884**	**0.236**	**1.273E-03**	**0.699**	**0.905**	**0.288**	**1.282E-03**	**0.595**
Arvengarten	Western	CH-019	*0.840*	*0.220*	*1.182E-03*	*0.652*	*0.875*	*0.276*	*1.227E-03*	*0.570*
Bergün	Eastern	CH-023	0.864	**0.229**	**1.234E-03**	0.680	0.896	**0.286**	**1.276E-03**	0.590
Ritom	Eastern	CH-028	0.853	**0.229**	**1.244E-03**	0.673	0.890	**0.284**	**1.280E-03**	0.585
Sex Carro	Western	CH-032	0.859	0.225	1.213E-03	0.674	0.893	0.282	1.255E-03	0.584
Val Medel	Eastern	CH-034	0.862	0.226	1.222E-03	0.681	**0.898**	0.281	1.253E-03	**0.594**
Lago Sfii	Western	CH-035	0.859	**0.229**	**1.232E-03**	0.677	0.888	**0.284**	**1.270E-03**	0.584
Selva Secca	Eastern	CH-039	0.848	0.226	1.216E-03	0.667	*0.879*	0.280	1.249E-03	*0.575*
Uerlicherblase	Western	CH-045	*0.839*	*0.219*	*1.176E-03*	*0.656*	*0.878*	*0.275*	*1.231E-03*	*0.573*
Fafleralp	Western	CH-046	0.855	0.222	1.198E-03	0.669	0.889	0.280	1.245E-03	0.582
Meder	Western	CH-052	*0.821*	*0.216*	*1.164E-03*	*0.632*	*0.866*	*0.270*	*1.215E-03*	*0.564*
Untersteinberg	Western	CH-053	*0.841*	0.224	1.208E-03	*0.658*	0.884	0.281	1.262E-03	0.581
Bürchen	Western	CH-113	**0.889**	0.226	**1.228E-03**	**0.706**	**0.913**	**0.284**	**1.269E-03**	**0.603**
God Giavagl	Eastern	CH-150	**0.887**	**0.234**	**1.266E-03**	**0.709**	**0.910**	**0.287**	**1.283E-03**	**0.603**
Davos	Eastern	EC-HJ	**0.880**	**0.229**	**1.244E-03**	**0.696**	**0.906**	**0.285**	**1.279E-03**	**0.599**
Rautialp	Eastern	EN-HJ	*0.811*	*0.209*	*1.129E-03*	*0.610*	*0.850*	*0.262*	*1.173E-03*	*0.543*
Celerina	Eastern	ES-HJ	**0.912**	**0.235**	**1.276E-03**	**0.730**	**0.931**	**0.291**	**1.303E-03**	**0.619**
Grengiols	Western	WC-HJ	*0.837*	*0.218*	*1.178E-03*	*0.650*	*0.880*	*0.277*	*1.237E-03*	*0.574*
Kandersteg	Western	WN-HJ	0.861	0.222	1.204E-03	0.672	0.890	0.278	1.250E-03	0.581
Zermatt	Western	WS-HJ	**0.869**	*0.221*	*1.192E-03*	0.682	**0.901**	0.278	*1.238E-03*	0.589
Riederalp	Western	WZ-HJ	0.865	0.223	1.213E-03	**0.685**	0.890	0.280	1.254E-03	0.587
		Average	0.857	0.224	1.211E-03	0.672	0.890	0.280	1.253E-03	0.584
		Minimum	0.811	0.209	1.129E-03	0.610	0.850	0.262	1.173E-03	0.543
		Maximum	0.912	0.236	1.276E-03	0.731	0.931	0.291	1.303E-03	0.619

TABLE 3 Correlation between geographic peripherality (GP), habitat suitability (HS) and genetic diversity at neutral and adaptive loci using ten different SNP sets (Table 1) and four genetic diversity estimates (Table 2). The correlation coefficients *r* are based on Pearson's correlation, the *p* values on ANOVAs (significant models in bold).

			Geographic peripherality (GP)		**Habitat suitability (HS)**	
SNP set	**Criterion**	**Index**	***r***	***p* value**	***r***	***p* value**
Full SNP set	All SNPs (SNP_all)	PPL	-0.527	**0.009**	0.222	0.292
		H_e	-0.461	**0.025**	-0.059	0.789
		π	-0.460	**0.026**	-0.017	0.940
		Θ_W	-0.506	**0.013**	0.158	0.460
Neutral SNP sets	Tajima's *D* (SNP_neutral_D)	PPL	-0.520	**0.011**	0.198	0.355
		H_e	-0.439	**0.035**	0.016	0.941
		π	-0.437	**0.036**	0.012	0.956
		Θ_W	-0.508	**0.013**	0.150	0.488
	π variation (SNP_neutral_pi)	PPL	-0.530	**0.009**	0.216	0.308
		H_e	-0.465	**0.025**	-0.041	0.852
		π	-0.478	**0.021**	0.011	0.960
		Θ_W	-0.510	**0.013**	0.161	0.453
	F_{ST} outliers (SNP_neutral_XTX)	PPL	-0.528	**0.009**	0.222	0.293
		H_e	-0.467	**0.023**	-0.060	0.786
		π	-0.477	**0.020**	-0.017	0.937
		Θ_W	-0.508	**0.013**	0.162	0.450
	EAA (SNP_neutral_LFMM)	PPL	-0.528	**0.009**	0.227	0.281
		H_e	-0.465	**0.024**	-0.061	0.783
		π	-0.458	**0.027**	-0.017	0.938
		Θ_W	-0.501	**0.014**	0.154	0.472
	Overall (SNP_neutral_overall)	PPL	-0.514	**0.012**	0.218	0.306
		H_e	-0.442	**0.034**	0.030	0.890
		π	-0.459	**0.027**	0.013	0.954
		Θ_W	-0.500	**0.014**	0.143	0.507
Adaptive SNP sets	Tajima's *D* (SNP_adaptive_D)	PPL	-0.298	0.165	0.140	0.502
		H_e	0.078	0.722	-0.724	**< 0.001**
		π	-0.070	0.750	-0.308	0.144
		Θ_W	-0.192	0.373	0.266	0.208
	π variation (SNP_adaptive_pi)	PPL	-0.458	**0.024**	0.257	0.201
		H_e	-0.368	0.055*	-0.144	0.501
		π	-0.337	0.088	-0.114	0.599
		Θ_W	-0.436	**0.033**	0.125	0.556
	F_{ST} outliers (SNP_adaptive_XTX)	PPL	-0.119	0.590	0.147	0.501
		H_e	-0.081	0.714	-0.005	0.982

	π	-0.106	0.621	-0.001	0.996
	Θ_W	-0.404	0.051	0.050	0.816
EAA (SNP_adaptive_LFMM)	PPL	-0.232	0.285	0.030	0.893
	H_e	0.029	0.893	0.015	0.945
	π	-0.195	0.370	-0.002	0.993
	Θ_W	-0.413	0.050	0.158	0.470

* p value of the model with quadratic term

1004

Figures

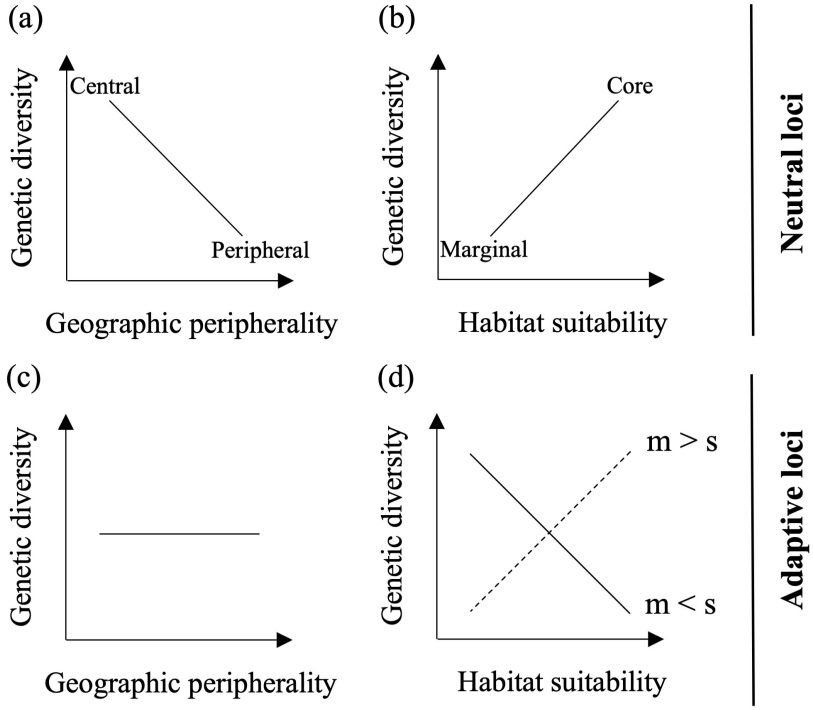


FIGURE 1 Hypothesised relationships between geographic peripherality (GP, a and c), habitat suitability (HS, b and d), and genetic diversity at neutral (a-b) and adaptive loci (c-d). For adaptive loci in relation to habitat suitability (d), two scenarios are presented; (i) with migration rate m > selection coefficient s (dashed line), and (ii) with m < s (solid line).

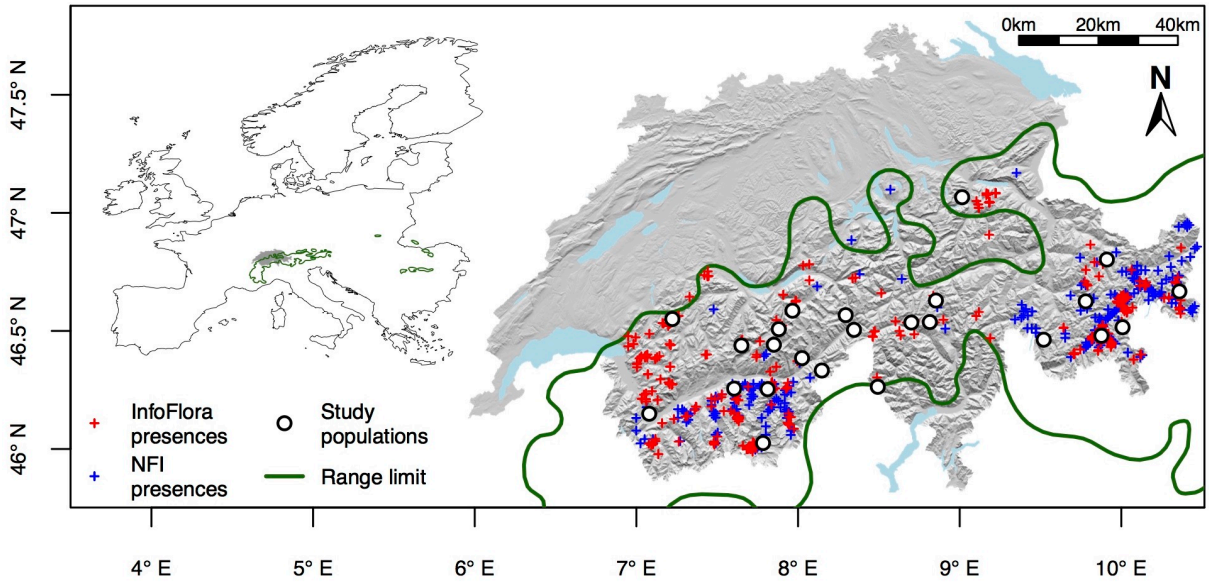
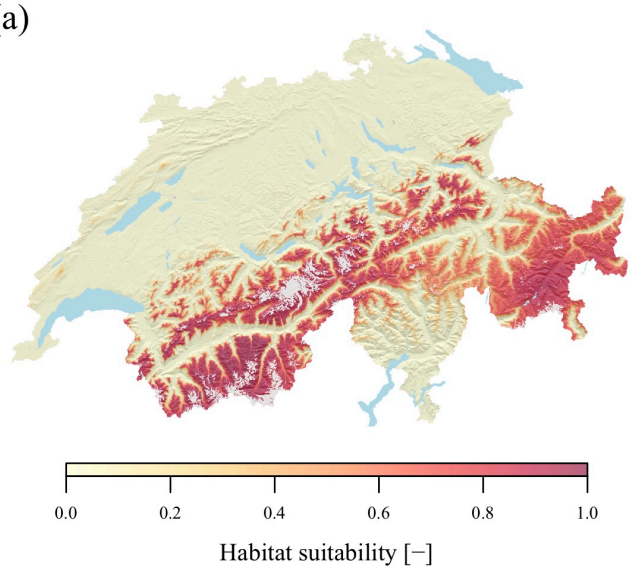


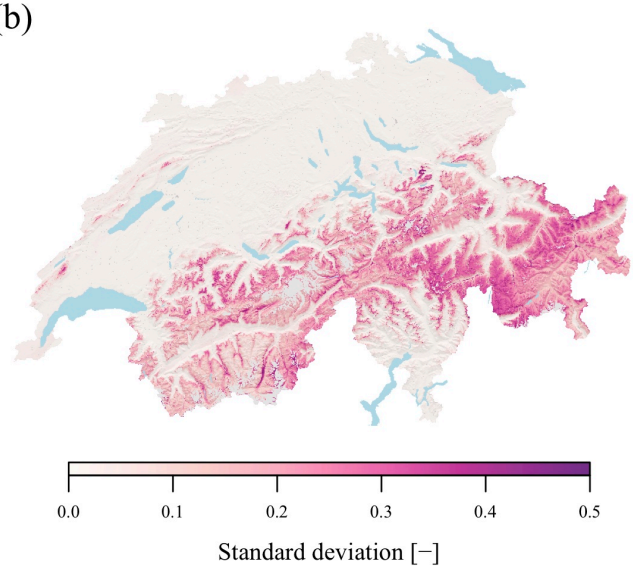
FIGURE 2 Natural range with occurrences and sampling sites of *Pinus cembra* in the Swiss Alps. A digital elevation model for Switzerland is used as background map (www.swisstopo.admin.ch), the range limit is derived from Caudullo et al. (2017). The inserted European map shows the study area and the complete geographical distribution of the species.

1017

(a)


0.0
0.2
0.4
0.6
0.8
1.0
Habitat suitability [−]

(b)


0.0
0.1
0.2
0.3
0.4
0.5
Standard deviation [−]

(c)

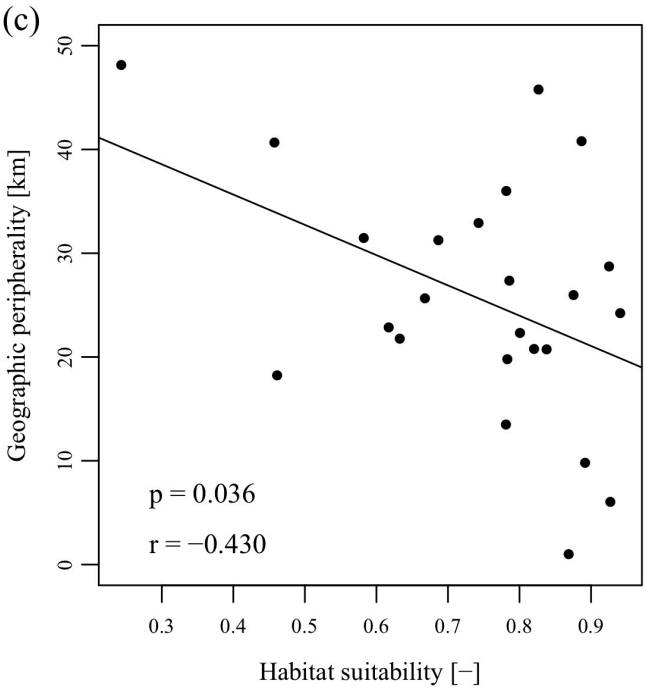

Geographic peripherality [km]
0
10
20
30
40
50
p = 0.036
r = −0.430
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Habitat suitability [−]

FIGURE 3 Predicted distribution of habitat suitability (HS) and its correlation with geographic peripherality (GP). (a) Weighted average of five species distribution models (SDMs) used for prediction (generalised linear model, generalised additive model, random forest, artificial neural network, and maximum-entropy). The values 0 and 1 mean the worst and the best environmental conditions for the studied species, respectively. (b) Standard deviation of the five SDMs. (c) Correlation between GP and HS with Pearson's correlation coefficients *r* and *p* value.

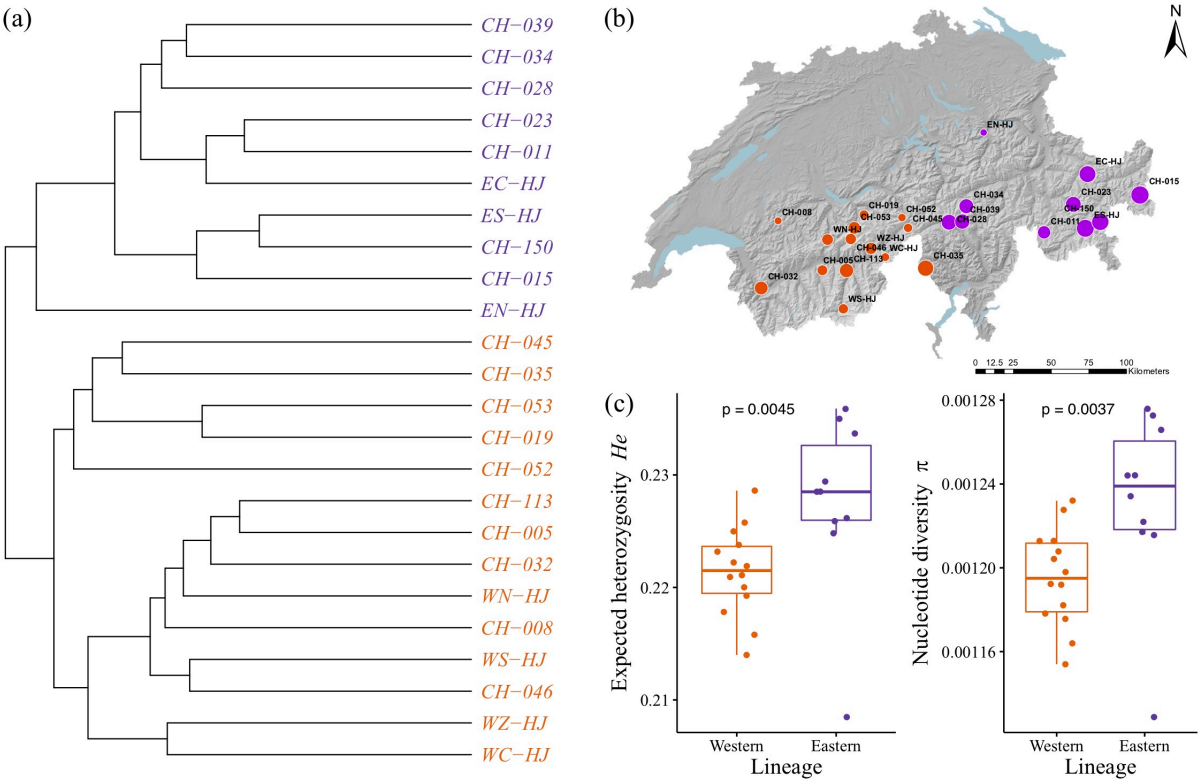


FIGURE 4 Genetic diversity and population structure in *Pinus cembra* across the species' Swiss range using the full SNP set (SNP_all). (a) Population structure based on a hierarchical clustering tree of Ω, with colours referring to the two main phylogeographic lineages. (b) Map of the studied populations, with colouring of population codes as in (a) and circle sizes denoting expected heterozygosity H_e. (c) Boxplots showing differences in genetic diversity (H_e and nucleotide diversity π) between eastern and western lineages.

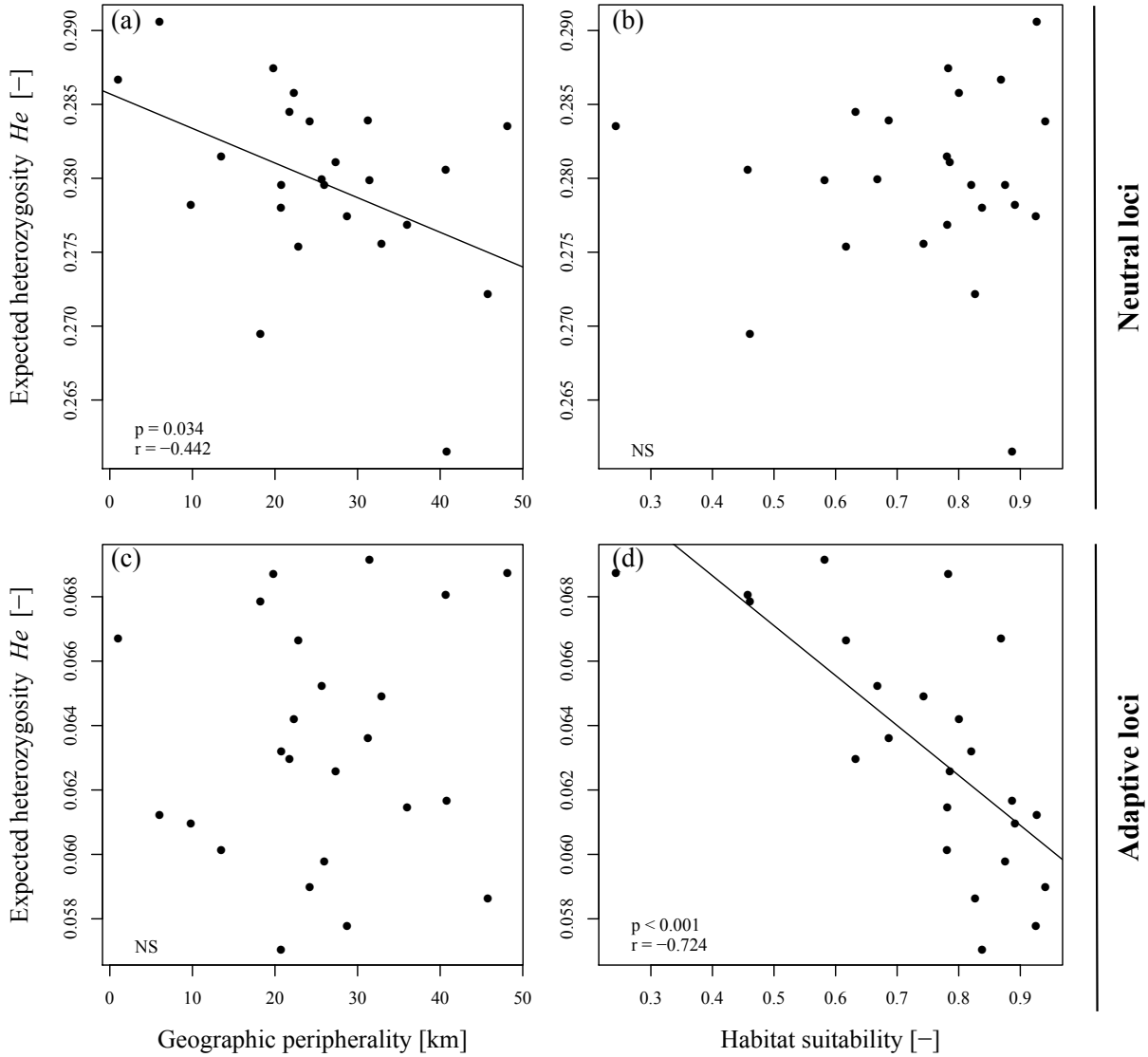


FIGURE 5 Correlation between geographic peripherality (GP, a and c), habitat suitability (HS, b and d), and expected heterozygosity (H_e) at putatively neutral (a-b) and adaptive loci (c-d) in *Pinus cembra*. The neutral SNP set presented in (a) and (b) consisted of SNPs that were not identified as putatively adaptive in any of the four sets of adaptive loci. The adaptive SNP set (c-d) is based on Tajima's *D*.