Genomic variation of a keystone forest tree species reveals signals of local adaptation despite high levels of phenotypic plasticity

Desanka Lazic[1], Cornelia Geßner[1], Katharina J. Liepe[1], Isabelle Lesur-Kupin[2], Malte Mader[1], Céline Blanc-Jolivet[1], Dušan Gömöry[3], Mirko Liesebach[1], Santiago C. González-Martínez[2], Matthias Fladung[1], Bernd Degen[1], Niels A. Müller[1]*

1 Thünen Institute of Forest Genetics, Grosshansdorf, Germany.
2 BIOGECO, INRAE, University of Bordeaux, Cestas, France.
3 Faculty of Forestry, Technical University in Zvolen, Zvolen, Slovakia.

*Correspondence to NAM (niels.mueller@thuenen.de)

Local adaptation is key for ecotypic differentiation and species evolution. Understanding the underlying genomic patterns can allow the prediction of future maladaptation and ecosystem stability. Here, we report the whole-genome resequencing of 874 individuals from 100 range-wide populations of European beech (*Fagus sylvatica* L.), one of the most important forest tree species in Europe. We show that genetic variation closely mirrors geography with a clear pattern of isolation-by-distance. Genome-wide analyses for genotype-environment associations (GEAs) identified relatively few potentially adaptive variants after correcting for an overwhelming signal of statistically significant but non-causal GEAs. We characterized the single high confidence genomic region and pinpoint a candidate gene possibly involved in winter temperature adaptation acting by modulating spring phenology. Surprisingly, however, allelic variation at this locus did not result in any apparent fitness differences in a common garden. More generally, reciprocal transplant experiments across large climate distances demonstrated extensive phenotypic plasticity. Nevertheless, we find indications of polygenic adaptation which may be essential in natural ecosystems. This polygenic signal exhibits broad- and fine-scale variation across the landscape highlighting the relevance of spatial resolution. In summary, our results emphasize the importance but also exemplify the complexity of employing natural genetic variation for forest conservation under climate change.

Keywords: Local adaptation, adaptive genetic variation, genotype-environment associations, missing heritability, phenotypic plasticity, genomic offset, *Fagus sylvatica*.

Introduction

Terrestrial plants and specifically trees make up the majority of the Earth's biomass[1,2], thereby giving forests an essential role for global carbon fluxes. With storage potentials reaching up to 25% of the current atmospheric carbon pool, forest restoration might represent one of the most effective strategies for climate change mitigation[3,4]. However, rather than being restored to mitigate climate change, forests are themselves severely threatened by rapidly changing climates and may experience marked habitat reductions[5,6]. To better understand and counteract this threat, there is an urgent need for reliable estimates of species' responses to climate change. Such estimates can be derived from species distribution models, which are based on current climate limits. However, these models treat species as homogenous and static units and do not account for adaptive differentiation and the variable adaptive capacity of local populations[7,8]. The increasing feasibility of generating large population genomic data to elucidate the genetic basis of local adaptation, and combining these data with reciprocal transplant and common garden experiments calls for an integration of intraspecific genetic variation into conservation biology frameworks[9-11].

To elucidate the genetic basis of local adaptation, one straightforward approach is to perform genotype-environment association (GEA) analyses. For this, all that is needed in principle is genomic

data for a sufficiently large number of georeferenced individuals adapted to different environments, although additional factors, such as confounding population structure or the expected genetic complexity, should be considered[12,13]. By associating each sequence variant in the genome with the relevant environmental variables, loci potentially involved in adaptation to differences in those variables can be identified. These data can then be integrated into prediction models of species' responses to climate change. The usefulness of GEAs for such predictions has already been demonstrated for conservation biology, e.g., for predicting vulnerability of coral reefs to heatwaves[14].

Additionally, genomic and environmental information can be integrated to estimate 'genomic offset', which is a proxy for potential future maladaptation and has been empirically demonstrated to outperform climate distance models[15,16]. Despite inherent uncertainties[17-19], such as the impact of different training sets[20] or the specific relationship between local adaptation and future climate vulnerability[21], genomic offset estimates could contribute to guiding future breeding and conservation efforts. This has been exemplified for animals and plants, such as the yellow warbler, a migratory songbird in North America[22], or pearl millet, a cereal crop in West Africa[23]. But perhaps the most striking possible application for genomic offset estimates can be envisioned for long-lived organisms such as trees. Due to their long juvenile phase and exceedingly long generation times, the fitness of populations cannot easily be experimentally tested under different environmental conditions. If, however, the natural sequence variants associated with adaptation to specific environmental conditions were known, future performance of populations could be predicted based on the gap between the current and the required future genetic make-up. Populations that appear perfectly healthy today may exhibit markedly different levels of maladaptation in the future. It should be noted though that this not only requires the presence of (pre)adaptive genetic differentiation between populations but also a precise knowledge of the environmental drivers of potential future population decline.

In Central Europe, the ecologically and economically most important broadleaf species is European beech (*Fagus sylvatica* L.), henceforward referred to as beech. It represents a prime example of a long-lived species likely affected by future climate maladaptation. Currently, beech still represents the dominant species of the potential natural vegetation (PNV) in many European countries[24]. Its wide distribution range extends from Spain in the southwest, to Sweden in the north all the way to Greece and Bulgaria in the southeast, highlighting its huge ecological amplitude. Accordingly, evidence of local adaptation has been reported at different geographical scales[25,26], although phenotypic plasticity appears pervasive[27,28]. Tree ring data indicate that beech may experience growth rate declines of 20-50% in 70 years from now[8]. However, these predictions do not consider natural genetic variation and thus ignore possible differences in the adaptive potential of different local populations. To gain a detailed understanding of the genetic architecture of local adaptation and potential patterns of maladaptation under future climates, we aimed to characterize genomic variation of beech across its distribution range and relate it to tree performance in two common gardens.

Results

Genomic variation mirrors geography in European beech

To sample the range-wide genetic diversity of beech, we took advantage of a common garden being comprised of 100 populations from across the range (Supplementary Fig. 1, Supplementary Table 1). We randomly selected nine trees per population for whole-genome resequencing with an average sequencing depth of 42.5x (Supplementary Table 2). Mapping the resequencing data to the chromosome-level beech reference genome[29], we identified a total of 3.68 million high-confidence

sequence variants, that is SNPs and short indels, of which about 540 thousand were largely independent, exhibiting linkage disequilibrium (LD) values below 0.2. Analysis of these independent variants revealed the presence of close relatives, such as half sibs or first cousins, in some populations (Supplementary Fig. 2). The material used for the establishment of the common garden thus only involved a limited number of seed trees in some cases. To avoid any artefacts in our subsequent analyses due to kinship structure, we removed individuals with pairwise relatedness of 2nd degree or higher. Additionally, we excluded one genetically highly divergent population from Bulgaria, potentially representing a hybrid with the sister species *F. orientalis* (Lipsky), and one apparently admixed population from Germany (Supplementary Fig. 3). This left us with a final set of 653 individuals from 98 populations (Supplementary Table 3).

Strikingly, a two-dimensional visualization of the genomic variation of these 653 individuals by principal component analysis (PCA) revealed a close correspondence with the map of Europe (Fig. 1a). Especially individuals from the western part of the range, that is Spain, France, Germany, Denmark and Sweden, exhibited a remarkable correlation between the first two principal components (PCs) and geography. The relationship between PC1 and longitude was high across all populations with an adjusted R^2 of 0.91 (Fig. 1b,d). PC2, however, varied mostly by latitude (adjusted R^2=0.66, Fig. 1c,e). Marked geographical structure can be observed up to PC6 (Supplementary Fig. 4).

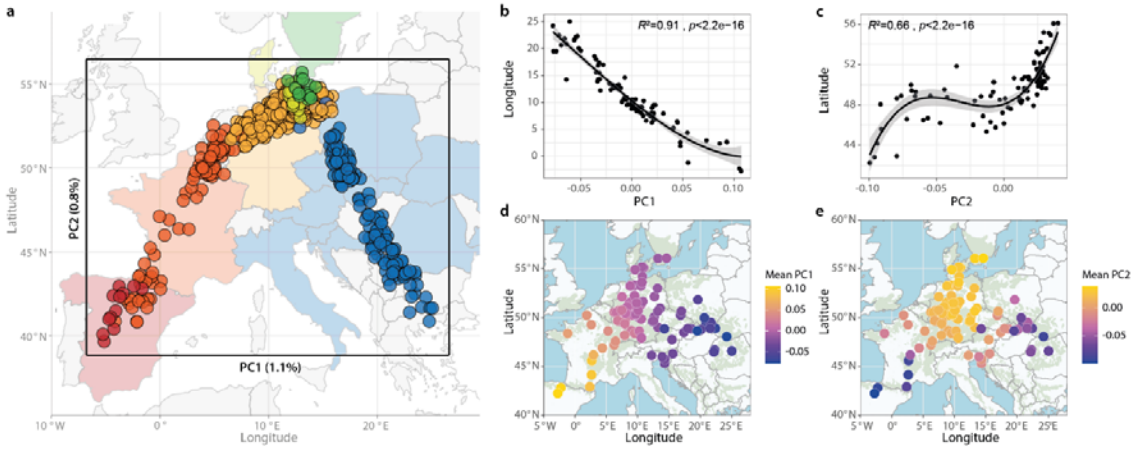


Figure 1: Genomic variation across the distribution range of European beech (*Fagus sylvatica*) mirrors geography. (a) Overlay of a map of Europe and the results of a principal component analysis (PCA) using 540k independent genome-wide variants in 653 largely unrelated (less than 2nd degree) individuals from 98 populations shows correlation between geography and PC1 and PC2, which explain 1.1% and 0.8% of the total genetic variation, respectively. Each point represents one of the 653 individuals plotted according to their PC1 and PC2 values. Colors correspond to the country of the origin of their source populations: Spain = red, France = orange, Germany = yellow, Denmark = light green, Sweden = dark green, other countries = blue. (**b**, **c**) Polynomial regression (3rd order) of average PC1 and longitude (b), and PC2 and latitude (c) reveals highly significant (p<2.2e-16) relationships (adjusted R^2=0.91 and 0.66, respectively). Each population is marked by a black point, bold lines and shading indicate regression and 95% confidence interval of the model. (**d**, **e**) Geographic origins (longitude and latitude) of the 98 analyzed populations are depicted by circles. Colors indicate the population means of PC1 (d) and PC2 (e). Green background shading represents the distribution range of beech.

Population genomics reveals three ancestral genetic clusters and isolation-by-distance

Analysis of the individuals' ancestry coefficients using 'sparse non-negative matrix factorization' (snmf)[30] indicated the presence of three ancestral genetic clusters and substantial admixture (Fig. 2a, Supplementary Fig. 5). The cluster with the highest number of individuals in our dataset dominates the central part of the distribution range while the other two ancestry clusters are mainly found in the east and in the west, respectively (Fig. 2a,b). Whereas the genetic diversity is similar among

clusters (π=1.66e-3 to 1.73e-3) genetic differentiation is highest between the more distant eastern and western groups with an F_{ST} value of 0.040 compared to 0.029 (central vs. west) and 0.016 (central vs. east) (Fig. 2c). Despite the generally low levels of genetic differentiation, the populations exhibit a highly significant signal of isolation-by-distance (Mantel test r=0.79, p<0.001) which becomes especially apparent due to the large geographical extent of our sampling (Fig. 2d).

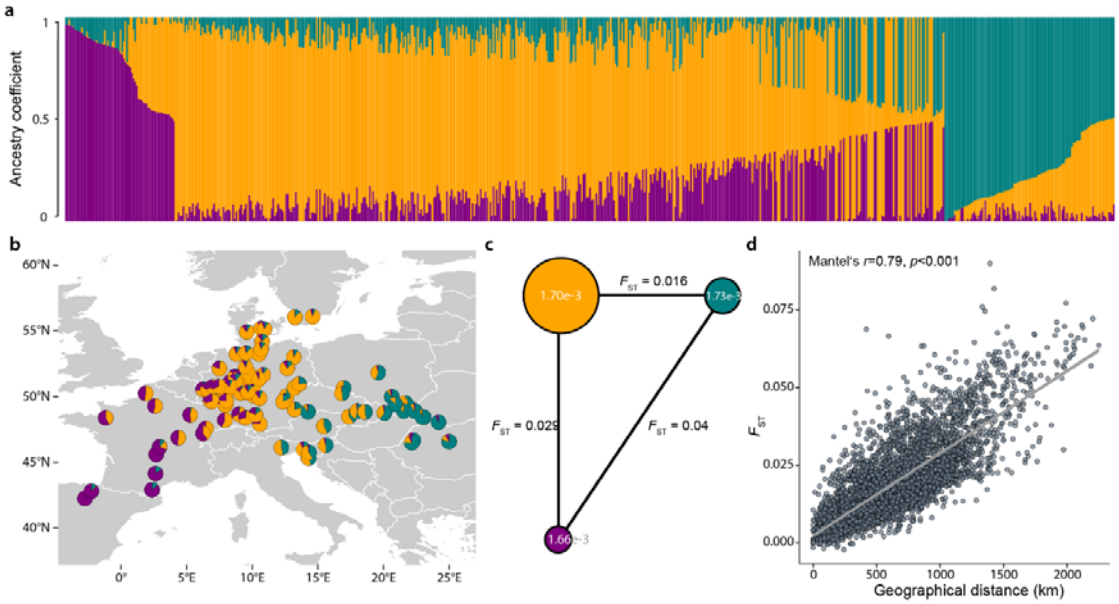


Figure 2: Population genomics reveals three ancestral genetic clusters and significant isolation-by-distance. (**a**) Analysis of the individuals' ancestry coefficients indicated the presence of three major ancestral genetic clusters (K=3) whose proportions are represented for each of the 653 unrelated beech individuals by purple, orange and green bars, sorted by cluster membership. (**b**) Most populations belong to the central cluster shown in orange, while the clusters shown in green and purple dominate in the East and in the West, respectively. (**c**) Genetic diversity (π) indicated by white numbers within the circles is similar among the three clusters, which are depicted as circles scaled by the number of individuals with more than 70% cluster membership. Genetic differentiation (F_{ST}) shown by the length and the numbers on the lines increases with geographic distance. (**d**) Pairwise genetic differentiation (F_{ST}) between populations compared to geographic distances reveals significant (Mantel's r=0.79, p<0.001) isolation-by-distance. Grey line depicts corresponding linear model.

Together, these results demonstrate a high level of accuracy of the common garden used for our analyses, as any mistakes in the organization or planting of the seeds and seedlings would have resulted in outliers or an overall erosion of the geographic signal in the genomic data. Additionally, the results show the absence of any pronounced human impact by seed transfer on the genetic composition of our populations. All populations, except a single one from the southern Czech Republic (showing as an outlier in Fig. 1d,e or Fig. 2b), appear largely autochthonous and can therefore be considered representative of natural populations that have evolved under the environmental conditions of their origins. Considering the large environmental amplitude of the populations sampled, relatively strong selection for adaptive differentiation may be expected. The autochthony of the populations together with the broad environmental range open the exciting possibility of identifying genotype-environment associations (GEAs) that can provide insights into the genetic basis of local adaptation.

<u>Genotype-environment association (GEA) analyses suffer from pervasive random signal</u>

To explore the genetic basis of local adaptation, we extracted the 19 bioclimatic variables from the WorldClim database[31]. These interpolated climatic indices provide a powerful resource to assess

155 GEAs, although other environmental data types and sources may yield additional insights[13]. As
156 expected from the geographic origins of our populations, the bioclimatic variables show broad
157 mostly normal distributions (Supplementary Fig. 6). Non-clinal genetic patterns arising from natural
158 selection under certain scenarios can limit the fraction of locally adaptive sequence variants that can
159 be accurately inferred[32]. Nevertheless, well-powered landscape genomics datasets may still enable
160 the detection of adaptive clinal variation. In line with this, our data exhibited thousands of
161 statistically significant (Bonferroni correction, p<9.25e-8) GEAs scattered across the genome in a
162 latent factor mixed model (LFMM) analysis[33]. However, performing the same analysis with
163 randomized geographical coordinates revealed a highly similar signal of statistically significant
164 associations, even after false discovery rate (FDR) or Bonferroni correction. Since the genotypes are
165 not associated with the random coordinates in any biologically meaningful way, this signal is non-
166 causal by definition (Supplementary Fig. 7). Importantly, this indicates that running a genotype-
167 environment association analysis and simply correcting for multiple testing may yield an exceedingly
168 high number of false positives, that is variants that are statistically associated but have no biological
169 relevance.

170 High levels of false positives, of close to 100%, have already been reported in simulation studies for
171 different GEA methods[32,34]. To identify the genetic basis of local adaptation, true signal thus needs to
172 be separated from the apparently overwhelming signal of random associations. We aimed to tackle
173 this challenge by comparing the p-value distribution of random GEA runs with that of our real data to
174 estimate the likelihood of a signal to occur by chance (Supplementary Fig. 7). While this, as any hard
175 threshold, cannot differentiate causal from non-causal associations and will thus inevitably lead to
176 false negatives (i.e., true associations not being identified), it provides a measure of confidence for
177 the potential adaptive variants in a specific dataset. Using this randomization approach to determine
178 a significance threshold for which the likelihood of being exceeded with random data compared to
179 the real data is 5% (number of associations with random data/number of associations with real data
180 = 1/20), only one single locus on chromosome 2, associated with winter temperature, remained (Fig.
181 3a, Supplementary Fig. 8). Lowering the threshold to 50%, a level at which half of the associations are
182 expected to occur by chance, eight additional loci were identified (Fig. 3a).

183 <u>A highly associated locus for winter cold may cause adaptive differentiation by modulating phenology</u>

184 The single high-confidence locus on chromosome 2 also stands out when using all 3.68 million
185 variants (Supplementary Fig. 9) or other GEA methods, such as BayPass[35] or WZA[34] (Fig. 3b-d).
186 However, with random data these methods pick up just as many associations as with the real data
187 (Supplementary Fig. 10). The same is true for the multivariate RDA method[36] which shows that a
188 substantial part of the explainable genetic variance in our data can be attributed to differences in the
189 environment (Supplementary Fig. 11, Supplementary Table 4). Isolation-by-environment (IBE) of
190 associated variants, corrected for geographic structure, sometimes used as an indicator of local
191 adaptation, was significant (partial Mantel test, p<0.001) for real and random data thus also not
192 serving to distinguish causal from non-causal signal (Supplementary Fig. 12). A more detailed view of
193 the high-confidence genomic region shows extended linkage disequilibrium (LD) across
194 approximately 300 kb (Fig. 3e). Thus, a natural sequence variant within this region is likely involved in
195 the adaptation to different winter temperatures.

196 The projection of the most significantly associated variant onto the map shows higher frequencies of
197 the alternative haplotype, associated with lower winter temperatures, in the eastern part of the
198 range but also in some high-altitude populations such as provenance 11 from central France
199 (Supplementary Fig. 1, Fig. 3f). At the same time, the map of haplotypes highlights that the
200 associated sequence variants represent common alleles that are not fixed and confined to specific
201 populations or geographic regions but are segregating across the entire distribution range. Notably,

the reference allele of the cold-associated haplotype dominantly delays spring phenology by about 2.5 days (one-way ANOVA and Tukey's *post hoc* HSD test, $p<0.001$, Fig. 3g) which could contribute to frost avoidance in the less predictable climates of the Atlantic part of the range. Interestingly, one annotated gene in the region, encoding a Callose synthase 1 (Bhaga_2.g94), has been shown to play an important role in controlling winter dormancy in European aspen (*Populus tremula* L.)[37]. This gene was highly expressed, albeit not differentially expressed between genotypes, in winter buds of eight beech trees sampled in December (Fig. 3h). It could thus differentially regulate winter dormancy and spring bud burst as an adaptation to different climates.

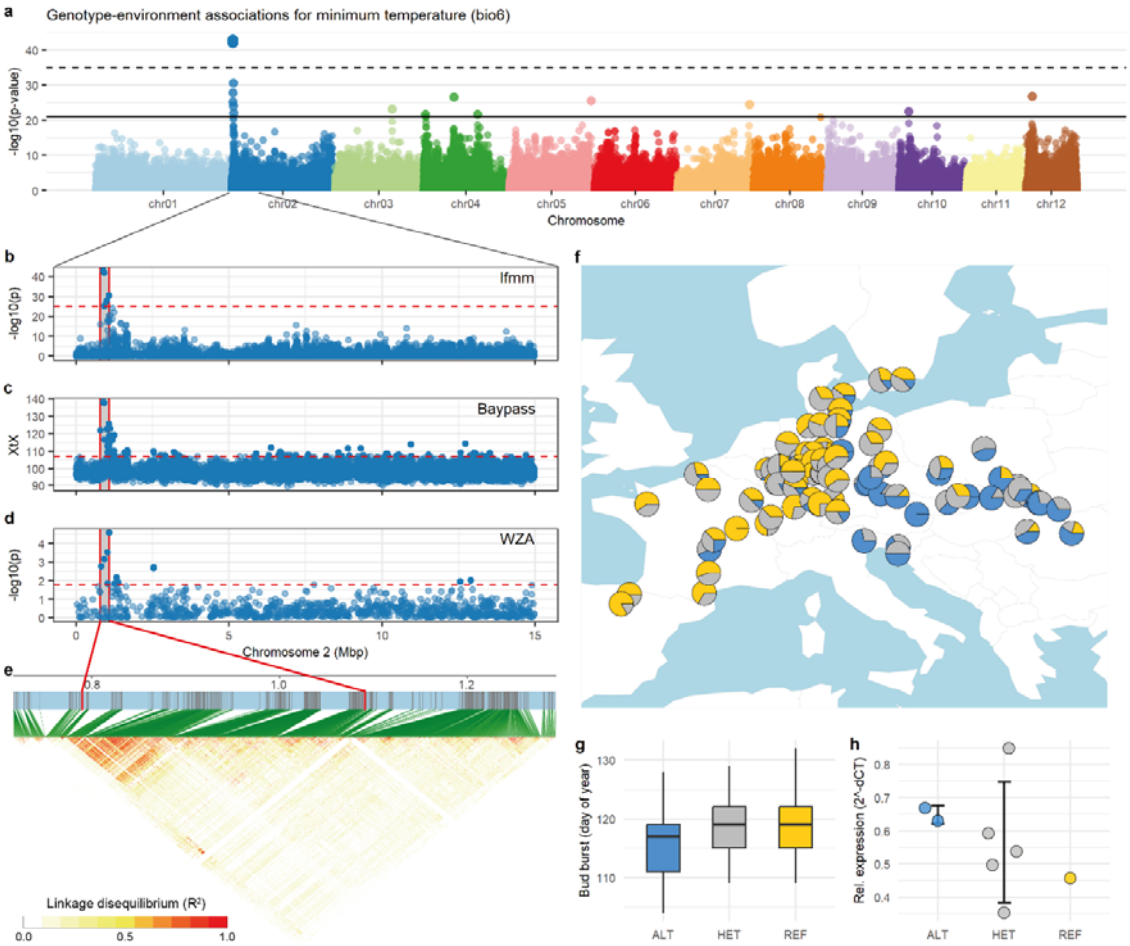
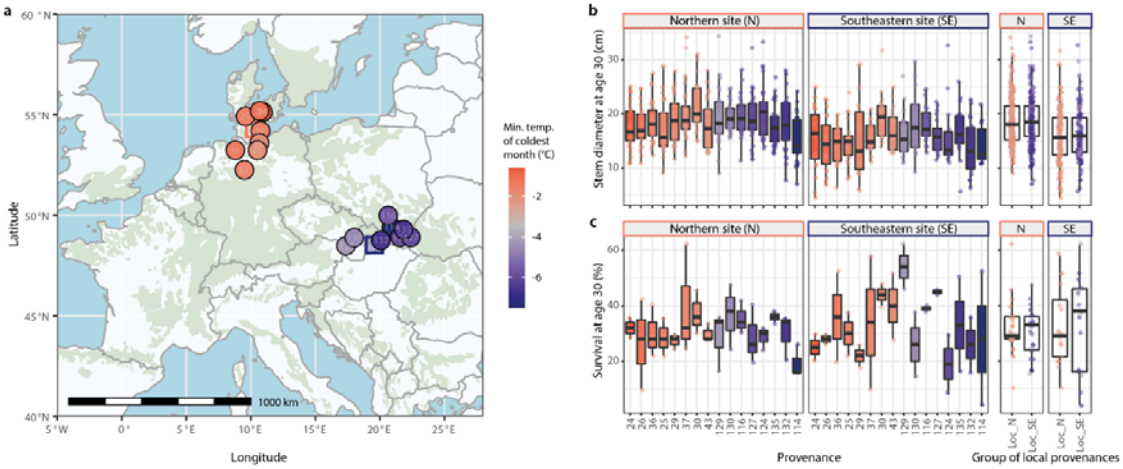


Figure 3: Genotype-environment associations reveal genomic signals of local adaptation. (**a**) Manhattan plot of the 540k LD-filtered genome-wide variants in 653 individuals from 98 populations. The significance of the association of each variant along the 12 beech chromosomes with the minimum temperature of the coldest month (bio6) of the population origins is indicated by the -log10 *p*-values on the y-axis. The dashed and solid horizontal lines indicate significance thresholds for which the likelihood with random data compared to the real data is 5% and 50%, respectively (Supplementary Fig. 7). (**b**, **c**, **d**) Close-up view of the lfmm results (b) and two alternative methods, BayPass XtX (c) and WZA z-scores (d), of the highest confidence genomic region on chromosome 2 delineates the locus to a 300 kb region (0.79-1.09 Mbp). (**e**) The pairwise linkage disequilibrium (R^2) for non-LD-filtered variants is shown. Red colors indicate high R^2 values. The associated genomic region is marked by red lines. Green lines connect LD values to physical positions on the chromosome. (**f**) Pie charts show geographic distribution of genotypes for the top lfmm hit (chr2:898,040); colors indicate genotypes homozygous for the reference allele (REF, yellow), heterozygous (HET, grey) or homozygous for the alternative allele (ALT, blue). Only populations with more than three individuals are depicted. (**g**) Boxplots show spring phenology, i.e. day of bud burst, for the unrelated beech individuals in the common garden ordered by their genotype for the same variant shown in (f) (ALT: n=157, HET: n=265, REF: n=229). The REF allele dominantly causes later bud burst (one-way ANOVA and Tukey's *post hoc*

225 HSD test, $p<0.001$). (**h**) Relative expression (2^{-dCT}) of the *CALLOSE SYNTHASE 1* gene in winter buds was determined for eight
226 trees using qRT-PCR. Circles represent relative expression, error bars ± 1 standard deviation.

227 Common garden and reciprocal transplant experiments demonstrate extensive phenotypic plasticity

228 The winter cold locus also shows up in a local analysis, which can increase statistical power[38], using
229 individuals from the central genetic cluster only (Supplementary Figs. 13 and 14). Given its strong
230 association across different geographic scales and its effect on potentially adaptive phenotypic
231 differentiation, one might expect it to have an effect on fitness. Trees carrying the locally adaptive
232 reference allele should show higher levels of fitness in our common garden in northern Germany
233 than trees with the alternative allele. In beech, one important proxy for fitness appears to be growth
234 with larger trees generating more offspring[39]. Surprisingly though, allelic variation of the winter cold
235 locus did not affect growth, that is stem circumference, of the beech trees in our common garden in
236 northern Germany (one-way ANOVA, p=0.51, Supplementary Fig. 15). One explanation for this could
237 be conditional neutrality, where the locus is beneficial only in one but neutral in the other climate. To
238 test this possibility, we performed a reciprocal transplant experiment by analyzing two common
239 gardens strongly differing for the minimum temperature of the coldest month and comprising
240 provenances from the two contrasting local climates (Fig. 4a). Strikingly, however, again we did not
241 find any indications of local adaptation when analyzing growth, i.e., stem diameter at breast height
242 (DBH), or survival at age 30 (Fig. 4b,c). While we identify provenance- and site-specific differences for
243 growth (but not survival), there is no difference between groups of local provenances nor a
244 significant genotype-environment interaction (two-way ANOVA, p=0.42 and p=0.51, respectively, Fig.
245 4b,c) which is considered a hallmark of local adaptation[40]. Given the consistent genetic
246 differentiation between these local populations (Supplementary Fig. 16) and the large climate
247 distance between the two common gardens, our results suggest extensive phenotypic plasticity.



248

249 **Figure 4: Reciprocal transplant experiment demonstrates extensive phenotypic plasticity.** (**a**) We selected two common
250 gardens, indicated by rectangles, strongly differing for the minimum temperature of the coldest month (bio6) and
251 comprising two contrasting sets of local provenances, eight from the Atlantic and eight from the continental climatic region,
252 depicted by circles. Colors indicate average minimum temperatures, grey marking the median value across the beech
253 distribution range, which is shown by green shading. (**b**,**c**) Boxplots show growth, i.e., stem diameter at breast height (DBH),
254 for the best 10 out of 50 originally planted trees in each of three blocks at the Northern (N) site (n=30) and two blocks at
255 the Southeastern (SE) site (n=20) (b) and survival among these 50 trees for each of the blocks (n=3 and n=2 in the N and the
256 SE site, respectively) (c), for eight Atlantic and eight continental provenances, individually and combined. Colors of boxes
257 and points follow the same color scheme as in (a).

258 Does this phenotypic plasticity in our common gardens thus imply the absence of local adaptation in
259 beech? And does it mean that any beech population could grow equally well anywhere across the
260 range? While this can be true for a single planted generation under specific environments, the
261 situation may markedly change under different environmental conditions, and especially in natural
262 forest ecosystems. With competition by other species, rare extreme weather events and natural
263 selection during the seedling stage, relevant performance and fitness differences that are masked in
264 our artificial common garden system could become visible. The signal identified in our genotype-
265 environment association analyses supports the importance of adaptive phenotypic differentiation in
266 natural beech ecosystems. We therefore wanted to further explore this differentiation beyond the
267 single high-confidence locus identified in our GEA analysis by characterizing the genetic architecture
268 of spring phenology in a genome-wide association study (GWAS) for bud burst.

269 Genome-wide association studies (GWAS) reveal polygenicity and missing heritability

270 Phenology is geographically differentiated across the distribution range of beech[41] and has been
271 indicated to be of adaptive relevance in various tree species[42,43]. In poplar, a single major locus has
272 been shown to control local adaptation for autumn phenology, i.e. the timing of bud set[44]. A single
273 locus affecting the timing of spring phenology also turned up in our GEA analysis. Does this indicate a
274 simple genetic architecture with a single major locus driving phenotypic differentiation in phenology
275 in beech as well? To address this question, we performed a GWAS for bud burst assessed during two
276 years in the same 653 beech individuals as used in the GEA analysis. While we can predict phenotypic
277 variation for bud burst with an accuracy of up to 0.45 using genomic prediction models
278 (Supplementary Fig. 17) we did not identify any consistent GWAS loci (Supplementary Fig. 18).
279 Significant associations changed between years, despite a high correlation of phenotypes (Pearson's
280 r=0.889, p<2.2e-16), and only explained a small fraction of the overall phenotypic variation (1.4 to
281 5.1%, Supplementary Fig. 18) demonstrating high levels of missing heritability. These results indicate
282 a complex genetic architecture with a large number of small-effect loci resulting in insufficient
283 statistical power for genetic dissection. Growth, which as mentioned before is a proxy for fitness in
284 beech[39], is genetically even more intractable, with virtually no explanatory power of genomic
285 prediction and no significant associations in our GWAS (Supplementary Figs. 17 and 18).

286 Importantly, GEA analyses are conceptionally similar to GWAS, only with unmeasured composite
287 fitness phenotypes approximated by the environment instead of direct phenotypic measurements.
288 Thus, it may not be surprising that the genetic basis of environmental adaptation cannot be resolved
289 with relatively small sample sizes of less than 1000 individuals either and that heritability remains
290 largely missing[45-47]. Taken together, we interpret our single high-confidence GEA locus as an
291 indication of high missing heritability caused by a polygenic genetic architecture. We find additional
292 indications of this polygenic signal in our GEA results. For some of the bioclimatic variables, e.g.,
293 isothermality (bio3), the ratio of real vs. random signal increases relatively steeply already for lower-
294 significance variants (Supplementary Fig. 19), suggesting the presence of many small-effect loci
295 below the formal significance threshold. The machine learning algorithm 'Gradient Forests', which
296 can be used to model turnover functions of allele frequencies in response to environmental
297 variables[48], also highlighted bio3 as the most important bioclimatic variable in our dataset
298 (Supplementary Fig. 20) even though no individual locus stood out (Supplementary Fig. 8). Just as
299 using our genomic data for reliable prediction of bud burst despite not having the power to dissect
300 the individual loci in a GWAS, we might be able to predict maladaptation to future environments
301 using 'genomic offset' models without knowing the underlying genetic basis.

302 Genomic offset reveals broad- and fine-scale patterns of potentially adaptive genetic variation

303 Even if genomic offset between current and future conditions, based on current genotype-
304 environment associations, may not always predict future performance due to compensating effects
305 of plastic responses or differences between local adaptation (local vs. foreign) and future
306 maladaptation (home vs. away)[21], it can still reveal interesting spatial patterns of potentially adaptive
307 genetic variation in relation to climate change. We therefore assessed genomic offset in our data by
308 employing two conceptionally different methods. First, we calculated 'risk of non-adaptedness'
309 (RONA) for isothermality (bio3), which as described above exhibits the strongest signal of polygenic
310 adaptation in our data, to estimate the distance between current and putatively required future
311 allele frequencies weighted by the strength of their linear associations[49,50]. Second, we used the
312 Gradient Forests (GF) method, to calculate the genomic offset using nonlinear turnover functions and
313 all 19 bioclimatic variables[48].

314 The two methods, which were significantly correlated (Pearson's r=0.3, p<0.01, Supplementary Fig.
315 21), both indicated higher offsets in the southwestern part of the range (Fig. 5). Due to the extensive
316 phenotypic plasticity in our common gardens only a small fraction of the phenotypic variance for
317 growth is explained by the populations' origins (one-way ANOVA, adjusted R^2=1.8%, Supplementary
318 21). We could therefore not validate our genomic offset estimates, even though RONA values
319 calculated with respect to our common garden instead of future climate in a space-for-time approach
320 exhibited a correlation with stem circumference (one-way ANOVA, adjusted R^2=0.9%, p<0.01,
321 Supplementary 21). One possibility to validate our genomic offset estimates in the future and to
322 evaluate the impact of different environmental variables may be the assessment of our populations
323 in their present natural habitats. Especially after recent extreme years, possible differences in current
324 maladaptation should become apparent.

325 For a more holistic view of species' responses to climate change, it was suggested to consider
326 additional variables such as gene flow, dispersal or genetic load[8]. Recessive genetic load was
327 independent of genomic offset in our data highlighting another layer of complexity (Supplementary
328 21). It also did not follow any particular geographical pattern suggesting that demographic history in
329 the sampling range did not involve extreme range expansions or colonization events affecting
330 effective population size (Supplementary Fig. 22), in agreement with the similar levels of nucleotide
331 diversity found across genetic clusters. Most importantly, our genomic offset analyses highlighted
332 geographically fine-scale variation irrespective of the exact climate change model or scenario (Fig. 5,
333 Supplementary Fig. 23). This is nicely exemplified by some populations from southern Germany,
334 where adjacent populations exhibit markedly different levels of genomic offset (Fig. 5,
335 Supplementary Fig. 24). This pattern appears to be only partially explained by the distribution of
336 predicted climate change (Supplementary Fig. 24) indicating the presence of fine-scale differences in
337 adaptive genetic variation between some populations and highlighting the importance of spatial
338 resolution.

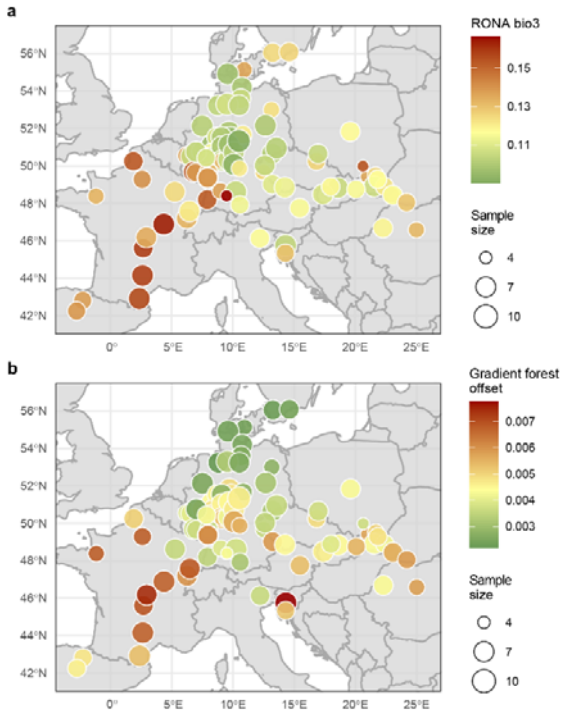


Figure 5: Genomic offset reveals broad- and fine-scale patterns of potentially adaptive genetic variation across the landscape. (**a**,**b**) The 'risk of non-adaptedness' (RONA) for isothermality (bio3) (a) and gradient forests offset for all 19 bioclimatic variables (b) were modelled using 3,321 associated (Bonferroni-adjusted $p<0.05$) and 10,000 randomly selected variants, respectively. Colors show estimated offset values, with yellow indicating the median. The size of the circles marks the number of individuals per population (n=4-10). Predicted climate in 2081-2100 for an intermediate climate change scenario (RCP4.5, MPI-ESM1-2-HR) was compared to near-current (1970-2000) conditions.

Discussion

The genomic analysis of 100 beech populations across the distribution range revealed a striking correlation between genetics and geography. The autochthonous populations may thus be locally adapted to different environmental conditions. Despite high levels of phenotypic plasticity under experimental common garden settings, we identified genomic signals of local adaptation. Importantly, our analyses emphasize the importance of data randomization and the need for larger sample sizes to achieve sufficient statistical power given the expected polygenic nature of adaptive traits. It will be exciting to see collaborative landscape genomics studies with a much higher number of sequenced trees. Only then can we start to fully appreciate the polygenic signal of adaptive differentiation, which will be important for further developing prediction models of forest growth and stability under climate change.

Our GEA and GWAS analyses suggest high levels of (potentially adaptive) standing genetic variation, which might contribute to adaptability. However, before any practical implementation, empirical validation of the relevance of this polygenic signal will be critical. Smartly designed new common garden experiments could contribute to this[21]. There should be an explicit focus on the specific effects of environmental variables. Under different environmental conditions, different environmental variables and thus different adaptive alleles will be most important for shaping the species adaptive responses, especially in the case of conditional neutrality. Additionally, future experiments should focus on more stressful sites similar to those expected under climate change. This might unmask so far hidden natural variation between populations. The possible relevance of

rare extreme weather events, competition and natural selection during the seedling stage should be considered for the experimental design. Also, different types of experiments may be needed depending on whether the research focus is silvicultural or ecological. Finally, for a reliable view of the adaptive genomic variation, spatial resolution is critical as a continuous distribution of adaptive alleles and interpolation may not be warranted[20]. Given that the selection of loci analyzed apparently does not play much of a role[15], this should become increasingly feasible using moderate density genotyping platforms. Large genetic biodiversity monitoring projects are being discussed and high-resolution sampling of different plant and animal species are on the way[51,52].

In conclusion, using genomics to predict future species performance and ecosystem stability is still a rapidly developing field. Especially in long-lived trees it promises great potential for informing forest management and conservation[53]. Careful variable selection and rigorous validation will be necessary before practical implementation. Our work using extensive genome-wide data across the range of a keystone forest tree species shows the importance but also the complexity of using genomics to resolve patterns of local adaptation and better understand climate change effects. Due to limited sample size and the polygenic architecture of adaptive traits, we are only skimming the surface of the encoded genomic information. Additionally, our common garden data point to the possibility of plasticity overlaying the expression of adaptive differentiation. Nevertheless, with increasing genomic and phenotypic data and new experimental sites, enhanced prediction models will likely contribute to conserving and restoring forests for biodiversity and as one of the most important carbon sinks for climate change mitigation.

Material and methods

Plant material and sampling. For the sample collection of beech genotypes from across the distribution range, we employed one of 23 common garden experimental sites of an international beech provenance trial planted in 1995. Specifically, we used the site near Schädtbek, Germany (trial code BU1901), where individuals from 100 different populations are growing in a randomized complete block design with multiple-tree plots of originally 50 individuals[54,55]. The coordinates of these populations (provenances) and the trial site are given in Supplementary Table 1. In each of the three blocks, three individuals were randomly selected from each population, totaling 900 samples. The only selection criterion was that the crown of the sampled trees is part of the canopy, to allow phenotyping with unmanned aerial vehicles (UAVs). Tree circumference at breast height was measured for all sequenced trees of each of the 100 populations in December 2022. Spring phenology was determined every two to three days from mid of April until mid of May 2022 and 2023 and defined as the majority of the tree crown showing leaves emerging from the winter buds. The phenotypic measurements are provided in Supplementary Table 5.

Reciprocal transplant experiment. Our study site in Schädtbek (BU1901) is part of the 4th international beech trial series sown in 1993 and planted in 1995[54]. A parallel site with a comparably large number of provenances is Vrchdobroč (BU1905) in Slovakia. Located at 840 m above sea level in the continental part of the distribution range, it experiences a contrasting climate with much lower minimum temperatures. This site was planted with two replications. During dormancy in 2022/23 a planned full assessment of 30-year diameter at breast height (DBH) was conducted at both sites. Here, we focused on eight provenances of local proximity to each trial site, 16 in total (northern provenances: 24, 25, 26, 29, 30, 36, 37 and 43; southeastern provenances: 114, 116, 124, 127, 129, 130, 132 and 135) to test for an effect of local adaptation. Analysis of variance (ANOVA) was conducted based on the following linear mixed model: $Y_{ijk} = \mu + P_i + E_j + (P \times E)_{ij} + B(E)_{jk} + e_{ijk}$, where Y_{ijk} is the phenotypic observation of a trait made for the ith provenance (P), grown at

411 the jth environment (E), located in the kth block (B) within environment E. $P \times E$ represents the
412 provenance by environment interaction, μ is the overall experimental mean, and e is the
413 experimental error (residual). Further, the model was applied with two provenance groups rather
414 than individual provenances, each aggregating one set of eight local provenances. Response variables
415 were DBH of the 10 strongest, most competitive trees per plot (Supplementary Table 6), and survival
416 percent per plot derived from the number of living trees relative to their number at establishment
417 (Supplementary Table 7).

418 **DNA extraction and Illumina resequencing.** Young leaves were sampled in the field in May and June
419 2021 and placed on ice in collection microtubes (QIAGEN, Hilden, Germany) in a 96-well format.
420 Samples were stored at -20°C until DNA extraction, following a previously described protocol[56]. DNA
421 sample QC, library preparation and sequencing were performed by Novogene Europe (Cambridge,
422 UK). Of the 900 samples, 26 failed sample QC, probably due to low amounts of starting material. For
423 the remaining 874 samples, Illumina sequencing libraries were prepared and an average of 22.98 Gb
424 of 150 bp paired-end data, which corresponds to approximately 42.5x coverage of the beech
425 reference genome[29], were generated on the Illumina Novaseq 6000 platform with an S4 flow cell.
426 Additional details on the sequencing statistics can be found in Supplementary Table 2.

427 **Variant calling.** Variant calling was performed using GATK version 4.0.5.1 following the best practices
428 for germline short variant discovery where possible[57]. After read filtering, which consisted of (1)
429 removing reads containing adapters (5' Adapter: 5'-
430 AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT-3', 3' Adapter: 5'-
431 GATCGGAAGAGCACACGTCTGAACTCCAGTCACGGATGACTATCTCGTATGCCGTCTTCTGCTTG-3'), (2)
432 removing reads containing more than 10% undetermined bases, and (3) removing reads of low
433 quality, that is reads with a Qscore<= 5 for more than 50% of the bases, the clean data were mapped
434 against the chromosome-level beech reference genome[29] using bwa-mem version 0.7.12 (ref. [58]) with
435 the following parameters: -k 32 -M -R. Duplicate reads were marked with Picard tool's (v2.26.2)
436 'MarkDuplicates' (http://broadinstitute.github.io/picard/). Using GATK's 'HaplotypeCaller', gVCF files
437 were generated which were combined by GATK's 'CombineGVCFs'. Finally, joint genotyping was
438 performed using GATK's 'GenotypeGVCFs' (ref. [59]).

439 **Variant filtering.** Hard variant filtering, instead of the recommended VQSR that requires a validated
440 set of variants as a truth or training set not available for beech, was based on variant quality scores,
441 variant coverage, missing data, linkage disequilibrium, minor allele frequencies and heterozygosity.
442 Specifically, we first filtered the SNP and indel data independently, considering the recommendations
443 from the technical documentation on "Hard-filtering germline short variants" on the GATK website.
444 We therefore removed variants based on strand bias (FisherStrand (FS) > 60 & StrandOddsRatio
445 (SOR) > 3) and mapping quality (RMSMappingQuality (MQ) < 40, MappingQualityRankSumTest
446 (MQRankSum) < -12.5). Based on the distribution of the variant confidence score QualByDepth (QD)
447 we choose a more stringent cutoff of QD > 10, to remove any low-confidence variants. Filtering was
448 performed with bcftools v1.7 (ref. [60]).

449 We then extracted the variant sequencing depth values (DP) using vcftools v0.1.16 (ref. [61]) to
450 visualize the DP distributions. To avoid any potentially hemizygous or duplicated sequence variants
451 not resolved in the reference genome, we chose relatively stringent coverage cutoffs based on the
452 mode of the distribution. We only allowed for DP values from 25% below to 50% above the mode of
453 the distribution, which represent values in between haploid and diploid or diploid and tetraploid
454 coverage, respectively. This resulted in cutoffs from 24.9 to 49.8 for SNPs and from 24.8 to 49.7 for
455 indels. Additionally, we filtered out variants with more than 10% missing data and only kept biallelic
456 variants. This resulted in a final dataset of 11.97 million variants for 874 individuals.

457 Using PLINK v1.9 (ref. [62]) we first checked for general patterns in our dataset using a principle
458 component analysis (--pca). This analysis highlighted nine individuals, all from the easternmost
459 provenance from Bulgaria, as outliers (Supplementary Fig. 3). Since the origin of the Bulgarian
460 provenance (provenance 158) is close to the distribution range of the second beech species in
461 Europe, *Fagus orientalis* (https://www.euforgen.org/species/fagus-orientalis), these individuals may
462 not represent pure *Fagus sylvatica* and were therefore removed from the dataset. Additionally, we
463 excluded one population from Northern Germany that appeared admixed (provenance 32) and one
464 individual from a German population (individual B9) that was closely related to individuals from a
465 population from Slovakia (provenance 135) and may represent a planting error (Supplementary Fig.
466 3).

467 We then filtered for minor allele frequency of 0.01 and pruned variants in complete linkage
468 disequilibrium (LD) in windows of 100 variants (--indep-pairwise 100 25 0.99). We also removed
469 variants with extreme deviation from Hardy-Weinberg equilibrium (p < 1e-08) [63]. Finally, we checked
470 for individuals with excessive levels of missing data or heterozygosity, as previously described[64,65].
471 We removed a single individual with high levels of missing data (78% vs less than 1.6% for all other
472 individuals). Additionally, we removed eight individuals based on high levels of heterozygosity,
473 exceeding three standard deviations from the mean.

474 To identify related individuals, we used the program KING v2.3.0 (ref. [66]) with 540 thousand mostly
475 independent variants (LD < 0.2 in windows of 50 variants). This analysis identified 192 individuals
476 with a relatedness of second degree (first cousins) or higher, which we removed using the function '--
477 unrelated -degree 2' Our final dataset comprised 3.68 million variants, of which 540,566 variants
478 exhibit LD values below 0.2 in windows of 50 variants, for 653 "unrelated" individuals.

479 **Population genomics.** Employing the R package LEA v3.10.1 (ref. [67]), we imported our final 3.68
480 million total variants and 540,566 independent variants using the function 'ped2lfmm'. Using the
481 independent variants, population genetic structure was evaluated with a principal component
482 analysis and an admixture analysis that are both implemented in LEA and called by the pca()- and
483 snmf()-functions, respectively. Both methods indicated three genetic clusters in our data. Thus, we
484 determined ancestry coefficients with snmf using K=3 (Supplementary Table 8) and assigned our
485 individuals into the three ancestral genetic clusters, determined by a 70% cluster membership
486 probability threshold. Population differentiation (Weir and Cockerham's F_{ST}) between and nucleotide
487 diversity (π) within these clusters were calculated in 10kb non-overlapping sliding windows using
488 vcftools v.0.1.16 (ref. [61]). It should be noted that π generated using vcftools will have underestimated
489 values due to the treatment of all missing sites as monomorphic. We then created an F_{ST} matrix and
490 the geographic distance (km) matrix between the populations for the isolation-by-distance (IBD)
491 analysis. Mantel tests were run using the R package vegan v.2.4-6 (ref. [68]), with the significance being
492 estimated based on 999 permutations.

493 **Genotype-environment association (GEA) analyses.** For the analysis of genotype-environment
494 associations we first extracted 19 bioclimatic variables representing historical climate data for the
495 reference period from 1970-2000 from the WorldClim database[31] (v2.1) with a resolution of 5x5 km
496 for all 98 populations using the R packages geodata[69] and terra[70]. This resolution approximately
497 matches the precision of the geographic coordinates of our populations. As described above, we
498 used the R package LEA v3.10.1 (ref. [67]) to import the two final sets of variants (3.68M and 540k) with
499 the function 'ped2lfmm'. As our populations probably exhibit three main ancestral genetic clusters,
500 we chose K=3 latent factors to account for confounding effects caused by population structure.
501 Missing genotype data were imputed using the impute()-function in LEA. We fitted the latent factor
502 mixed model (LFMM) using the lfmm()-function in LEA with 5 repetitions, 10,000 iterations and a

503 burnin of 5,000. The bioclim variables were used as environmental file. *P*-values were extracted using
504 the lfmm.pvalues()-function and a conservative Bonferroni significance threshold was applied.

505 Additionally, we performed three randomizations of the geographic coordinates of the 98
506 populations to randomize the environmental variables but maintain population structure. We used
507 those random environmental data to re-run the LFMM analysis (with 3 repetitions, 6,000 iterations
508 and a burn in of 3,000). *P*-value distributions of the three random runs were compared with the *p*-
509 value distribution from the run using the real data to determine a significance threshold for which
510 the likelihood of being exceeded with random data compared to the real data is 5% (number of
511 associations with random data/number of associations with real data = 1/20) or 50% (Supplementary
512 Fig. 7). Finally, we ran a "local" LFMM analysis with individuals from the central genetic cluster only.
513 For this, we selected populations with at least three individuals with more than 70% membership to
514 the central genetic cluster (Supplementary Fig. 13, orange). Again, we also ran the local LFMM
515 analysis with randomized environmental data.

516 Further, we applied BayPass v2.4 (ref. [35]) and WZA (ref. [34]) as two additional methods for calculating
517 GEAs. Within BayPass we performed three replicate runs of the core model to estimate XtX and
518 population covariance matrix and then took the median values across independent runs. Using the
519 median covariate matrix, we ran three repetitions of the auxiliary covariate model and kept the
520 median values of Bayes Factor (BF) across replicate runs. To get the threshold values we constructed
521 null distributions of the XtX values, by performing simulations of pseudo-observed data (POD) using
522 the simulate.baypass() function. Significant threshold for BF was 10 (i.e., Jeffrey's rule for "strong
523 evidence"). For WZA we calculated allele frequencies using vcftools v.0.1.16 freq2 option[61].
524 Population allele frequencies were correlated with 19 bioclimatic variables using Kendall's τ statistic
525 in R. Genome-wide τ results were analyzed in 10 kb non-overlapping windows using the weighted-Z
526 analysis (WZA)[34]. Both BayPass and WZA analysis were repeated with randomized population
527 coordinates and thus randomized environments for the bio6 environmental variable.

528 We additionally applied the multivariate method redundancy analysis (RDA)[36] to calculate GEAs. This
529 analysis was performed using the R package vegan v.2.4-6 (ref. [68]). To avoid overfitting, we selected
530 12 variables, removing highly correlated ones (such as "maximum temperature of warmest month"
531 and "mean temperature of warmest quarter", or "precipitation of wettest month" and "precipitation
532 of wettest quarter") keeping the more specific variables, i.e., daily or monthly rather than quarterly
533 data. For further analysis, these variables were standardized (i.e., subtracted the mean and divided
534 by standard deviation). We first applied partial RDA (pRDA) to dissect the effects of climate,
535 population structure, and geography using four models (Supplementary Table 4). We used 12
536 bioclimatic variables representing climate ('clim' in our model), population scores along the first six
537 PCA axes performed on 540,566 variants as a proxy for population structure ('struct' in our model)
538 and longitude and latitude to represent geography ('geog' in our model). Population allele
539 frequencies were used as a response variable in all models. We then used pRDA to identify candidate
540 loci associated with the environment while correcting for population structure. We defined outliers
541 based on the previously described rdadapt function[71]. The same analysis was repeated for three sets
542 of randomized population coordinates and, thus, randomized environments.

543 Calculation of R^2 and visualization of LD pattern was done using LDBlockShow v.1.40 (ref. [72]). For
544 expression analysis of the Callose synthase 1 gene (Bhaga_2.g94), winter buds of eight beech trees
545 were collected in December 2022. Total RNA was extracted with the Spectrum Plant Total RNA kit
546 (Sigma-Aldrich, USA) according to the manufacturer's manual, Protocol A. Following that, DNase I
547 digestion was performed using the Turbo DNA-free kit (Invitrogen, USA). For cDNA synthesis, 2 µg of
548 RNA, Oligo (dT) primers and SuperScript IV reverse transcriptase (Invitrogen, USA) were used

549 following the manufacturer's protocol, using 10 µl reactions without RNaseOUT. Reverse
550 transcriptase quantitative PCR (qRT-PCR) was carried out in triplicates on a CFX96 Touch Real Time
551 PCR Detection System (Bio-Rad Laboratories GmbH, USA) using the SsoAdvanced Universal SYBR
552 Green Supermix (Bio-Rad Industries, Inc., USA) and a two-step PCR program with annealing
553 temperature of 60 °C. Relative expression levels were calculated using the $2^{-\Delta Ct}$ method[73]. The
554 primers are given in Supplementary Table 9.

555 **Genomic prediction and genome-wide association studies (GWAS).** For genomic prediction, we used
556 the kin.blup() function from the R package rrBLUP v4.6.3 (ref. [74]). For computing the additive
557 relationship matrix with the A.mat()-function, we randomly selected 30,000 sequence variants from
558 the final set of 540k LD-filtered variants. For each size of training population (n=150, 250, 350, 450
559 and 550) we ran 200 cross validations by randomly assigning the according number of individuals,
560 from the final set of 653 unrelated individuals, to the training and the validation populations.
561 Predictive accuracies were calculated using the Pearson's correlation between predicted and
562 measured phenotypic values. Phenotypic values included bud burst determined in two years and
563 stem circumference measured in one year as described above and given in Supplementary Table 5.
564 To identify genotype-phenotype associations, we performed GWAS using the R package GAPIT
565 v.3.1.0 (ref. [75]) using the BLINK model[76]. As a phenotypic input, we again used bud burst assessed
566 during two years (2022 and 2023) and stem circumference measured in one year in the same final
567 653 individuals. A Bonferroni multiple test threshold was used to determine significance.

568 **Isolation-by-environment (IBE).** Using vcftools v.0.1.16 (ref. [61]) we calculated pairwise F_{ST} between
569 98 populations separately for significant (q<0.01) and non-significant variants (q>=0.01), as identified
570 in LFMM analyses with real and random data. We generated the F_{ST}, the geographic distance (km),
571 and environmental distance matrices for both data sets. The environmental distance was calculated
572 as Euclidean distance for the bio6 environmental variable. Partial Mantel tests were run using the R
573 package vegan v.2.4-6 (ref. [68]), with the significance being estimated based on 999 permutations. The
574 same analysis was repeated with a randomized data set.

575 **Genetic load analyses.** Population genetic load was evaluated by computing the average recessive
576 and additive genetic load per individual, obtained by counting the average number of derived
577 mutations with deleterious effects in homozygosis or total, respectively, relative to synonymous
578 variants to adjust for heterozygosity differences across individuals (as in ref. [77]). We first polarized
579 variants using the *Q. robur* genome[78] and retained 284,046 SNPs for which we were able to
580 determine the ancestral state. Second, a custom database was constructed to perform genomic
581 annotations (see below) and validated by comparing predicted protein and CDS sequences with
582 those obtained from the Bhaga European beech reference genome[29]. Finally, SNP variants were
583 annotated for their predicted effects on known genes (e.g., amino acid changes or loss of function
584 mutations) using SnpEff[79] and used for the counts of deleterious and neutral-effect (synonymous)
585 variants (based on 57,308 annotations).

586 **Genomic offset analyses.** To estimate genomic offset of our studied populations, we employed the
587 'risk of non-adaptedness' (RONA) and 'Gradient Forests' offset measures[48,49]. Present (1970-2000)
588 and future (2081–2100) bioclimatic variables for every population and for our common garden in
589 Schädtbek (54°3' "N 10°28' "E) were extracted from the WorldClim database using the CMIP6 'MPI-
590 ESM1-2-HR', the 'BCC-CSM2-MR' and the 'EC-Earth3-Veg' models[80-82] with a moderate (RCP 4.5) and
591 a more pessimistic (RCP 8.5) shared socioeconomic pathway (SSCP) with a 2.5-minutes resolution
592 (5x5 km). For RONA we used the bioclimatic variable isothermality (bio3). We calculated the
593 population-specific allele frequencies of significant variants (Bonferroni-adjusted p<0.05), as
594 determined by the LFMM analysis, using the R package data.table[83] on the imputed genotype matrix
595 as used for the LFMM analysis. RONA values were calculated as the distance between the current

allele frequencies and the required frequencies under projected future climate conditions based on linear regressions (allele.freq~environmental.variable) weighted by the adjusted R^2 values of the linear models[50]. We randomly subsampled 10,000 SNPs for gradient forests (GF) analysis and removed six populations with less than five individuals. Using population allele frequencies and 19 bioclimatic variables, GF was conducted in R with 'gradientForest' function[84]. Following code previously described[53], we calculated GF predictions that rescaled near-current and future bioclimatic variables into common units of genomic turnover. Genomic offset was then calculated as Euclidean distance between near-current and future genomic composition.

Statistics and data visualization. For all statistical analyses and data visualization we used R v4.2.2 (ref. [85]) and the ggplot2 package[86].

Data availability

All sequencing data are available from NCBI's SRA under the accession number PRJNA1005581.

Code availability

The code used for the analyses is available under:

https://github.com/desa-la/Beech_landscape_genomics

Acknowledgements

We would like to thank A. Eikhof, K. Groppe, S. Jencsik, S. Benischek and M. Spauszus for technical assistance in the lab and in the field. We are grateful to P. Eusemann and other members of the Thünen Institute of Forest Genetics, and K. Budde and other members of the AForGeN group for helpful comments and discussions. This work was funded by a core grant from the Thünen Institute and a grant from the FNR Forest Climate Fund (2219WK60A4). Support from the French government in the framework of the IdEx Bordeaux University 'Investments for the Future' program / GPR Bordeaux Plant Sciences is also acknowledged.

Author contributions

Conceptualization: MF, BD, NAM. Data curation: DL, CG, KJL, DG, NAM. Formal analysis: DL, CG, KJL, MM, CB-J, I-LK, SG-M, NAM. Funding acquisition: BD, NAM. Investigation: DL, CG, NAM. Methodology: DL, CG, KJL, NAM. Supervision: DG, ML, SG-M MF, BD, NAM. Validation: CG, NAM. Visualization: DL, KJL, NAM. Writing – original draft: NAM. Writing – review & editing: All authors.

Competing interests statement

The authors declare that they have no competing interests.

References

1. Erb, K.H. *et al.* Unexpectedly large impact of forest management and grazing on global vegetation biomass. *Nature* **553**, 73-76 (2018).
2. Bar-On, Y.M., Phillips, R. & Milo, R. The biomass distribution on Earth. *Proc Natl Acad Sci U S A* **115**, 6506-6511 (2018).
3. Bastin, J.F. *et al.* The global tree restoration potential. *Science* **365**, 76-79 (2019).

632 4. Cook-Patton, S.C. *et al.* Mapping carbon accumulation potential from global natural forest
633 regrowth. *Nature* **585**, 545-550 (2020).
634 5. Dyderski, M.K., Paź, S., Frelich, L.E. & Jagodziński, A.M. How much does climate change
635 threaten European forest tree species distributions? *Global Change Biology* **24**, 1150-1163
636 (2018).
637 6. Martinez Del Castillo, E. *et al.* Climate-change-driven growth decline of European beech
638 forests. *Commun Biol* **5**, 163 (2022).
639 7. Exposito-Alonso, M. *et al.* Genomic basis and evolutionary potential for extreme drought
640 adaptation in *Arabidopsis thaliana*. *Nature Ecology & Evolution* **2**, 352-358 (2018).
641 8. Aguirre-Liguori, J.A., Ramírez-Barahona, S. & Gaut, B.S. The evolutionary genomics of species'
642 responses to climate change. *Nature Ecology & Evolution* **5**, 1350-1360 (2021).
643 9. Sork, V.L. *et al.* Putting the landscape into the genomics of trees: approaches for
644 understanding local adaptation and population responses to changing climate. *Tree Genetics*
645 *& Genomes* **9**, 901-911 (2013).
646 10. Flanagan, S.P., Forester, B.R., Latch, E.K., Aitken, S.N. & Hoban, S. Guidelines for planning
647 genomic assessment and monitoring of locally adaptive variation to inform species
648 conservation. *Evolutionary Applications* **11**, 1035-1052 (2018).
649 11. Waldvogel, A.-M. *et al.* Evolutionary genomics can improve prediction of species' responses
650 to climate change. *Evolution Letters* **4**, 4-18 (2020).
651 12. Lasky, J.R., Josephs, E.B. & Morris, G.P. Genotype–environment associations to reveal the
652 molecular basis of environmental adaptation. *The Plant Cell* **35**, 125-138 (2022).
653 13. Dauphin, B. *et al.* Re-thinking the environment in landscape genomics. *Trends in Ecology &*
654 *Evolution* **38**, 261-274 (2023).
655 14. Selmoni, O., Cleves, P.A. & Exposito-Alonso, M. Global coral genomic vulnerability explains
656 recent reef losses. *bioRxiv*, 2024.03.25.586253 (2024).
657 15. Fitzpatrick, M.C., Chhatre, V.E., Soolanayakanahally, R.Y. & Keller, S.R. Experimental support
658 for genomic prediction of climate maladaptation using the machine learning approach
659 Gradient Forests. *Mol Ecol Resour* **21**, 2749-2765 (2021).
660 16. Gain, C. *et al.* A Quantitative Theory for Genomic Offset Statistics. *Molecular Biology and*
661 *Evolution* **40**(2023).
662 17. Hoffmann, A.A., Weeks, A.R. & Sgro, C.M. Opportunities and challenges in assessing climate
663 change vulnerability through genomics. *Cell* **184**, 1420-1425 (2021).
664 18. Rellstab, C., Dauphin, B. & Exposito-Alonso, M. Prospects and limitations of genomic offset in
665 conservation management. *Evol Appl* **14**, 1202-1212 (2021).
666 19. Laruson, A.J., Fitzpatrick, M.C., Keller, S.R., Haller, B.C. & Lotterhos, K.E. Seeing the forest for
667 the trees: Assessing genetic offset predictions from gradient forest. *Evol Appl* **15**, 403-416
668 (2022).
669 20. Lind, B.M. *et al.* How useful are genomic data for predicting maladaptation to future climate?
670 *Glob Chang Biol* **30**, e17227 (2024).
671 21. Lotterhos, K.E. Interpretation issues with "genomic vulnerability" arise from conceptual
672 issues in local adaptation and maladaptation. *Evolution Letters* (2024).
673 22. Bay, R.A. *et al.* Genomic signals of selection predict climate-driven population declines in a
674 migratory bird. *Science* **359**, 83-86 (2018).
675 23. Rhone, B. *et al.* Pearl millet genomic vulnerability to climate change in West Africa highlights
676 the need for regional collaboration. *Nat Commun* **11**, 5274 (2020).
677 24. Hickler, T. *et al.* Projecting the future distribution of European potential natural vegetation
678 zones with a generalized, tree species-based dynamic vegetation model. *Global Ecology and*
679 *Biogeography* **21**, 50-63 (2012).
680 25. Kreyling, J. *et al.* Local adaptations to frost in marginal and central populations of the
681 dominant forest tree *Fagus sylvatica* L. as affected by temperature and extreme drought in
682 common garden experiments. *Ecology and Evolution* **4**, 594-605 (2014).

683 26. Gauzere, J., Klein, E.K., Brendel, O., Davi, H. & Oddou-Muratorio, S. Microgeographic
684 adaptation and the effect of pollen flow on the adaptive potential of a temperate tree
685 species. *New Phytologist* **227**, 641-653 (2020).
686 27. Gárate-Escamilla, H., Hampe, A., Vizcaíno-Palomar, N., Robson, T.M. & Benito Garzón, M.
687 Range-wide variation in local adaptation and phenotypic plasticity of fitness-related traits in
688 *Fagus sylvatica* and their implications under climate change. *Global Ecology and*
689 *Biogeography* **28**, 1336-1350 (2019).
690 28. Kurjak, D. *et al.* Inter-provenance variability and phenotypic plasticity of wood and leaf traits
691 related to hydraulic safety and efficiency in seven European beech (*Fagus sylvatica* L.)
692 provenances differing in yield. *Annals of Forest Science* **81**, 11 (2024).
693 29. Mishra, B. *et al.* A chromosome-level genome assembly of the European beech (*Fagus*
694 *sylvatica*) reveals anomalies for organelle DNA integration, repeat content and distribution of
695 SNPs. *Front Genet* **12**, 691058 (2021).
696 30. Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. & François, O. Fast and efficient estimation
697 of individual ancestry coefficients. *Genetics* **196**, 973-83 (2014).
698 31. Fick, S.E. & Hijmans, R.J. WorldClim 2: new 1-km spatial resolution climate surfaces for global
699 land areas. *International Journal of Climatology* **37**, 4302-4315 (2017).
700 32. Lotterhos, K.E. The paradox of adaptive trait clines with nonclinal patterns in the underlying
701 genes. *Proc Natl Acad Sci U S A* **120**, e2220313120 (2023).
702 33. Frichot, E., Schoville, S.D., Bouchard, G. & Francois, O. Testing for associations between loci
703 and environmental gradients using latent factor mixed models. *Mol Biol Evol* **30**, 1687-99
704 (2013).
705 34. Booker, T.R., Yeaman, S., Whiting, J.R. & Whitlock, M.C. The WZA: A window-based method
706 for characterizing genotype–environment associations. *Molecular Ecology Resources* **24**,
707 e13768 (2024).
708 35. Gautier, M. Genome-Wide Scan for Adaptive Divergence and Association with Population-
709 Specific Covariates. *Genetics* **201**, 1555-1579 (2015).
710 36. Capblancq, T. & Forester, B.R. Redundancy analysis: A Swiss Army Knife for landscape
711 genomics. *Methods in Ecology and Evolution* **12**, 2298-2309 (2021).
712 37. Tylewicz, S. *et al.* Photoperiodic control of seasonal growth is mediated by ABA acting on cell-
713 cell communication. *Science* **360**, 212-215 (2018).
714 38. Lopez-Arboleda, W.A., Reinert, S., Nordborg, M. & Korte, A. Global Genetic Heterogeneity in
715 Adaptive Traits. *Molecular Biology and Evolution* **38**, 4822-4831 (2021).
716 39. Westergren, M. *et al.* Low but significant evolutionary potential for growth, phenology and
717 reproduction traits in European beech. *Molecular Ecology* **n/a**(2024).
718 40. Kawecki, T.J. & Ebert, D. Conceptual issues in local adaptation. *Ecology Letters* **7**, 1225-1241
719 (2004).
720 41. Wuehlisch, G.v., Krusche, D. & Muhs, H. Variation in temperature sum requirement for
721 flushing of beech provenances. (1995).
722 42. Savolainen, O., Pyhäjärvi, T. & Knürr, T. Gene flow and local adaptation in trees. *Annual*
723 *Review of Ecology, Evolution, and Systematics* **38**, 595-619 (2007).
724 43. Alberto, F.J. *et al.* Potential for evolutionary responses to climate change – evidence from
725 tree populations. *Global Change Biology* **19**, 1645-1661 (2013).
726 44. Wang, J. *et al.* A major locus controls local adaptation and adaptive life history variation in a
727 perennial plant. *Genome Biology* **19**, 72 (2018).
728 45. Abdellaoui, A., Yengo, L., Verweij, K.J.H. & Visscher, P.M. 15 years of GWAS discovery:
729 Realizing the promise. *The American Journal of Human Genetics* **110**, 179-194 (2023).
730 46. Tiffin, P. & Ross-Ibarra, J. Advances and limits of using population genetics to understand
731 local adaptation. *Trends in Ecology & Evolution* **29**, 673-680 (2014).
732 47. Yengo, L. *et al.* A saturated map of common genetic variants associated with human height.
733 *Nature* **610**, 704-712 (2022).

48. Fitzpatrick, M.C. & Keller, S.R. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters* **18**, 1-16 (2015).
49. Rellstab, C. *et al.* Signatures of local adaptation in candidate genes of oaks (Quercus spp.) with respect to present and future climatic conditions. *Mol Ecol* **25**, 5907-5924 (2016).
50. Pina-Martins, F., Baptista, J., Pappas Jr, G. & Paulo, O.S. New insights into adaptation and population structure of cork oak using genotyping by sequencing. *Global Change Biology* **25**, 337-350 (2019).
51. Pearman, P.B. *et al.* Monitoring of species' genetic diversity in Europe varies greatly and overlooks potential climate change impacts. *Nature Ecology & Evolution* **8**, 267-281 (2024).
52. Hoban, S. *et al.* Global genetic diversity status and trends: towards a suite of Essential Biodiversity Variables (EBVs) for genetic composition. *Biological Reviews* **97**, 1511-1538 (2022).
53. Sang, Y. *et al.* Genomic insights into local adaptation and future climate-induced vulnerability of a keystone forest tree in East Asia. *Nat Commun* **13**, 6541 (2022).
54. Robson, T.M., Garzon, M.B. & BeechCOSTe52-database-consortium. Phenotypic trait variation measured on European genetic trials of Fagus sylvatica L. *Sci Data* **5**, 180149 (2018).
55. Liesebach, M. International beech provenance trial 1993/95 - site Schädtbek (Bu19-1). *Thünen Report* **62**, 131-138 (2017).
56. Bruegmann, T., Fladung, M. & Schroeder, H. Flexible DNA isolation procedure for different tree species as a convenient lab routine. *Silvae Genetica* **71**, 20-30 (2022).
57. Poplin, R. *et al.* Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* (2018).
58. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
59. van der Auwera, G. & O'Connor, B.D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*, (O'Reilly Media, Incorporated, 2020).
60. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-93 (2011).
61. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-8 (2011).
62. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
63. Ingvarsson, P.K. & Bernhardsson, C. Genome-wide signatures of environmental adaptation in European aspen (*Populus tremula*) under current and future climate conditions. *Evol Appl* **13**, 132-142 (2020).
64. Anderson, C.A. *et al.* Data quality control in genetic case-control association studies. *Nat Protoc* **5**, 1564-73 (2010).
65. Marees, A.T. *et al.* A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int J Methods Psychiatr Res* **27**, e1608 (2018).
66. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-2873 (2010).
67. Gain, C. & François, O. LEA 3: Factor models in population genetics and ecological genomics with R. *Mol Ecol Resour* **21**, 2738-2748 (2021).
68. Oksanen, J. *et al.* vegan: community ecology package. *R package* (2022).
69. Hijmans, R.J., Barbosa, M., Ghosh, A. & Mandel, A. geodata: download geographic data. *R package* (2023).
70. Hijmans, R.J., Bivand, R., Pebesma, E. & Sumner, M.D. terra: spatial data analysis. *R package* (2023).
71. Capblancq, T., Luu, K., Blum, M.G.B. & Bazin, E. Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources* **18**, 1223-1233 (2018).

72. Dong, S.-S. *et al.* LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Briefings in Bioinformatics* **22**(2020).
73. Schmittgen, T.D. & Livak, K.J. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* **3**, 1101-8 (2008).
74. Endelman, J.B. Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome* **4**(2011).
75. Wang, J. & Zhang, Z. GAPIT version 3: boosting power and accuracy for genomic association and prediction. *Genomics, Proteomics & Bioinformatics* **19**, 629-640 (2021).
76. Huang, M., Liu, X., Zhou, Y., Summers, R.M. & Zhang, Z. BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience* **8**(2018).
77. Gonzalez-Martinez, S.C., Ridout, K. & Pannell, J.R. Range Expansion Compromises Adaptive Evolution in an Outcrossing Plant. *Curr Biol* **27**, 2544-2551 e4 (2017).
78. Plomion, C. *et al.* Oak genome reveals facets of long lifespan. *Nat Plants* **4**, 440-452 (2018).
79. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**, 80-92 (2012).
80. Müller, W.A. *et al.* A higher-resolution version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR). *Journal of Advances in Modeling Earth Systems* **10**, 1383-1413 (2018).
81. Xin, X. *et al.* BCC BCC-CSM2MR model output prepared for CMIP6 CMIP. (Earth System Grid Federation, 2018).
82. Consortium, E.C.-E. EC-Earth-Consortium EC-Earth3-Veg model output prepared for CMIP6 ScenarioMIP. (Earth System Grid Federation, 2019).
83. Dowle, M. *et al.* data.table: extension of 'data.frame'. *R package* (2023).
84. Ellis, N., Smith, S.J. & Pitcher, C.R. Gradient forests: calculating importance gradients on physical predictors. *Ecology* **93**, 156-68 (2012).
85. R, C.T. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* (2022).
86. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*, (Springer-Verlag New York, 2016).