Genomic variation of a keystone forest tree species reveals patterns of local adaptation and future maladaptation

Niels A. Müller*, Cornelia Geßner, Malte Mader, Céline Blanc-Jolivet, Matthias Fladung, Bernd Degen

Thünen Institute of Forest Genetics, Grosshansdorf, Germany.

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

Local adaptation is key for ecotypic differentiation and species evolution[1]. Understanding the underlying genomic patterns allows the prediction of future maladaptation and ecosystem stability[2]. Here we report the whole-genome resequencing of 865 individuals from 100 range-wide populations of European beech (*Fagus sylvatica*), which is one of the most important forest tree species in Europe. We show that genetic variation closely mirrors geography. Adaptive variation identified by genotype-environment associations exhibits highly polygenic architectures, involving thousands of associated sequence variants across the genome. By modelling the 'genomic offset' of these sequence variants under projected future climate conditions, we identify broad- and fine-scale variation highlighting geographic regions as well as populations at potential elevated risk of mortality or local extinction. Our results emphasize the importance of considering natural genetic variation for forest conservation under climate change conditions.

Keywords: Population genomics, genotype-environment associations, local adaptation, genomic offset, Fagus sylvatica, adaptive variation

Introduction: Terrestrial plants and specifically trees make up the majority of the Earth's biomass[3,4], 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[5,6]. However, rather than being restored to mitigate climate change, forests are themselves severely threatened by rapidly changing climates and may experience marked habitat reductions[7,8]. To better understand and counteract this threat, there is an urgent need for reliable estimates of species' responses to climate change. Such estimates have classically been derived from species distribution models, which are based on current climate limits. However, these models do not account for dynamic population genetic processes. More and more studies and an increasing amount of genomic data call for an integration of natural genetic variation, local adaptation, selection and gene flow[9-15]. This integration of data is essential for precise predictions of future effects of climate change on keystone species and the corresponding ecosystems.

Initial empirical validations already demonstrated that an integration of genomic and environmental information to estimate 'genomic offset', that is the prediction of future maladaptation or climate vulnerability, outperforms naive climate distance models in different species[16,17]. Genomic offset estimates can thus guide breeding and conservation efforts. This has been exemplified for animals and plants, such as the yellow warbler, a migratory songbird in North America[18], or pearl millet, a cereal crop in West Africa[19]. 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 are precisely characterized, future performance of populations can 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.

In Central Europe, the ecologically and economically most important broadleaf species European beech (*Fagus sylvatica* L.), henceforward referred to as beech, represents a prime example of a long-lived species potentially affected by future climate maladaptation. Currently, beech still represents the dominant species of the potential natural vegetation (PNV) in many European countries[20]. Its wide distribution range extends from Spain in the southwest, to Sweden in the north all the way to Greece and Turkey in the southeast. However, its future might be less favorable, as growth rate declines of 20-50% are predicted by 2090, depending on the region and climate change scenario[8]. To better understand the genetic basis of local adaptation to different climates and predict future performance by integrating genetic information, we set out to characterize the range-wide genomic variation of beech. New developments in sequencing technologies make population genomics studies increasingly feasible in virtually any species of interest[21]. Despite some remaining methodological challenges[22-24], predictions of future maladaptation of keystone forest tree species and future stability of forest ecosystems can therefore now be fully explored.

Results and discussion: To sample the range-wide genetic diversity of beech and elucidate the genetic basis of local adaptation and potential future climate vulnerability, we took advantage of a common garden comprising trees from 100 populations from across the distribution 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 sequencing data to the chromosome-level beech reference genome[25], we identified a total of 3.68 million high-confidence sequence variants, that is SNPs and short indels, of which about half a million were largely independent, exhibiting linkage disequilibrium (LD) values below 0.2. Analysis of the 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 the sister species *F. orientalis*, 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 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 correlation between PC1 and longitude was high across all populations with a Pearson's correlation coefficient of -0.94 ($p < 2.2e-16$) (Fig. 1b, c). Marked geographical structure can be observed up to PC6 (Supplementary Fig. 4). These results demonstrate a high level of accuracy of the common garden used for our analyses and the absence of any pronounced human impact, especially by seed transfer, on the genetic composition of beech. All populations, except a single one from the Southern Czech Republic, appear largely autochthonous and can therefore be considered representative of the environmental conditions of their origins. Considering the large amplitude of environmental differences, relatively strong selection for adaptive differentiation can be expected. This opens the exciting possibility to identify genotype-environment associations (GEAs) and elucidate the genetic basis of local adaptation.

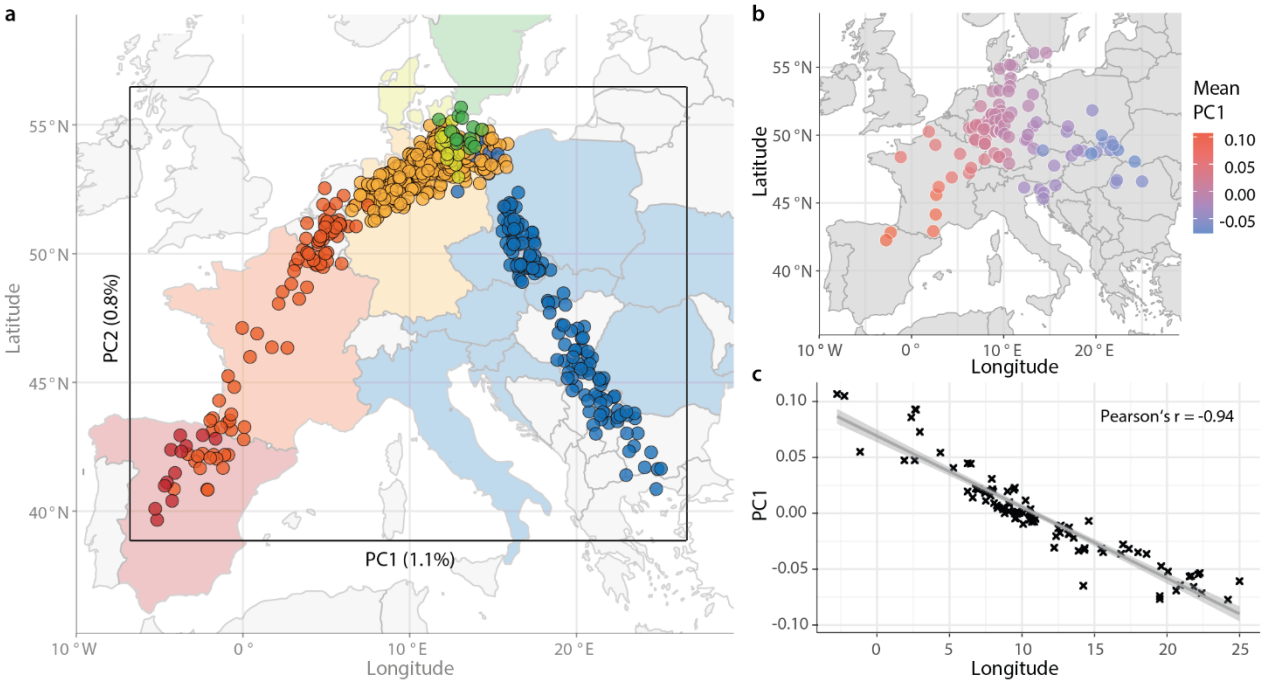

Figure 1: Genomic variation across the distribution range of European beech (*Fagus sylvatica*) mirrors geography. (**a**) Overlay of a map of Europe and the results of a principal component analysis (PCA) using 540k independent genome-wide variants in 653 largely unrelated (less than 2nd degree) individuals 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**) Geographic origins (longitude and latitude) of the 98 analyzed populations are depicted by circles. Colors indicate the population means of PC1. (**c**) Correlation analysis between PC1 and longitude reveals a Pearson's correlation coefficient *r* of -0.94 ($p < 2.2e-16$). Each population is marked by a black cross. Bold gray line and shading indicate linear regression and 95% confidence interval of the model, respectively.

To explore this possibility, we extracted 19 bioclimatic variables from the WorldClim database[26]. These interpolated monthly precipitation and temperature data provide a powerful resource to assess genotype-environment associations, although additional environmental data types and sources may further improve the corresponding analyses[27]. As expected from the geographic origins of our populations, the bioclimatic variables showed broad mostly normal distributions (Supplementary Fig. 5, Supplementary Table 4). Each variable exhibited thousands of significant GEAs scattered across the genome in a latent factor mixed model (LFMM) analysis[28] indicating a highly polygenic architecture of local adaptation (Fig. 2a, Supplementary Fig. 6).

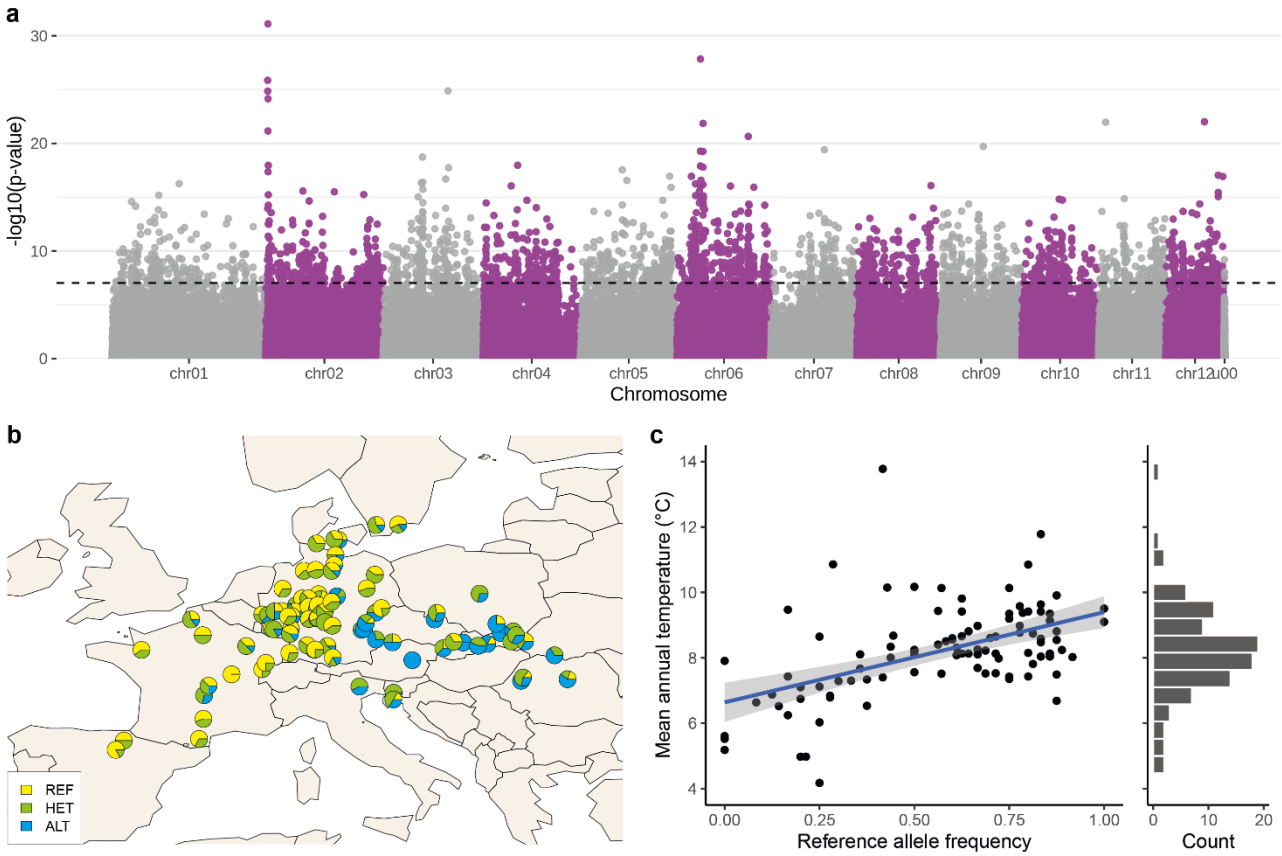

Figure 2: Genotype-environment associations reveal highly polygenic architecture. (**a**) Manhattan plot of 540k independent genome-wide variants in 653 individuals from 98 populations. The significance of the associations of each variant along the 12 beech chromosomes with the mean annual temperature (bio1) of the population origins is indicated by the -log10 p-values on the y-axis. In total, 2,758 variants exceed the Bonferroni threshold marked by the horizontal dashed line. (**b**) Geographic distribution of the genotypes for the sequence variant with the highest association, which is located at the beginning of chromosome 2, is shown by pie charts. Colors indicate genotypes: homozygous for the reference allele (REF, yellow), heterozygous (HET, green) or homozygous for the alternative allele (ALT, blue). Only populations with more than three individuals are depicted. (**c**) The frequency of the reference allele for the same variant shown in (b) and mean annual temperature of the population origins exhibit a significant linear association (adjusted R^2=0.25, p=1.205e-07). Each population is marked by a black point. Bold blue line and shading indicate linear regression and 95% confidence interval of the model, respectively. Histogram on the right shows distribution of mean annual temperature for the 98 populations.

Notably, theory predicts that such polygenicity makes natural selection most efficient[29]. In line with the emerging picture of polygenic adaptation rather than large-effect mutations and selective sweeps, putatively adaptive genetic variants in beech mostly represent common alleles that are not fixed and confined to specific populations or geographic regions but are segregating across the entire distribution range (Fig. 2b, c, Supplementary Fig. 7). The resulting high levels of potentially adaptive standing genetic variation suggest high levels of adaptability of beech populations to climate change in general. Nevertheless, allele frequency changes at many small-effect loci will be required for climate adaptation. Specific populations may thus differ in their risk of future maladaptation depending on their current genetic make-up. To formally test such differences across the landscape and potentially identify regions with elevated risk of mortality or even local extinction, we employed a genomic offset model called 'risk of non-adaptedness' (RONA). RONA predicts fitness under predicted future climate conditions based on distances between current allele frequencies of putatively adaptive variants (e.g., the 2,758 significantly associated variants for mean annual temperature in Fig. 2a) and required future allele frequencies. These distances are estimated from linear regressions weighted by the strength of their associations[30,31].

RONA estimates for two bioclimatic variables, that is mean annual temperature (bio1) and total annual precipitation (bio12), under an intermediate climate change scenario (representative concentration pathway, RCP 4.5) revealed marked variation between populations across Europe (Fig. 3). In general, RONA values, and thus potential future maladaptation, are higher towards the southern edges of the distribution range. This is consistent with predictions based on tree ring data[8] and indicates regions where future climate may exceed the species capacity of adaptation. Importantly, our analysis also highlighted geographically fine-scale variation as exemplified by some populations from Southern Germany for bio12, which exhibit more than twice the levels of genomic offset compared to adjacent populations (Fig. 3d). Overall, predicted RONA values, which can be interpreted as the allele frequency change required for adaptation to future climate conditions, are moderate. Allele frequency shifts in the range of 0.01 per generation were identified at neutral and adaptive loci in a different tree species[32]. According to our RONA estimates, shifts in that range could close the genomic offset gap and adapt most of the analyzed beech populations to average future climates (Fig. 3c, d). However, extreme weather conditions such as drought events could markedly change this picture. Also, when considering a more pessimistic climate change scenario (RCP 8.5), RONA values markedly increase for large parts of the distribution range and may put several populations at risk of local extinction (Supplementary Fig. 8).

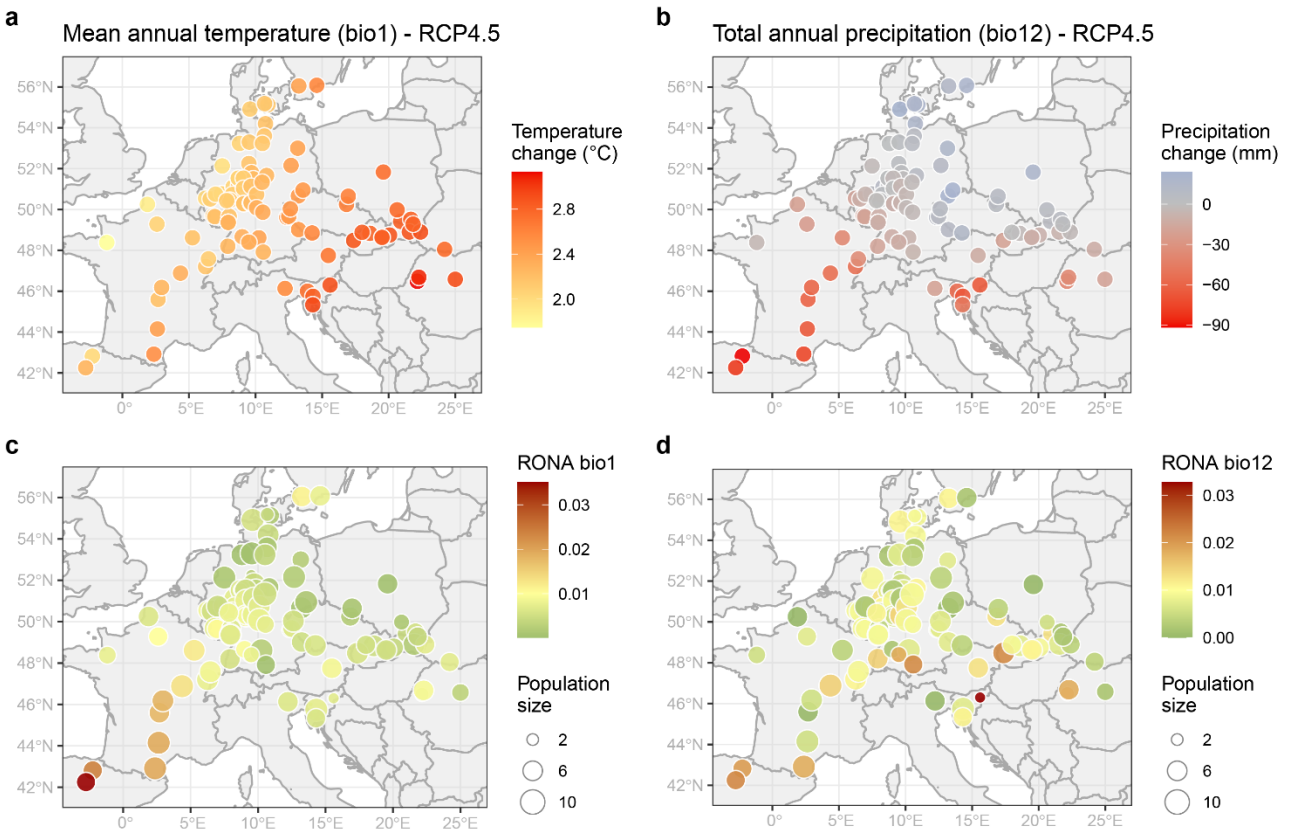

Figure 3: Prediction of future maladaptation reveals populations and regions with potential elevated risk of local extinction. (**a,b**) Predicted changes in mean annual temperature (bio1, °C) and total annual precipitation (bio12, mm) in 2081-2100 compared to near-current conditions are shown for the 98 analyzed beech populations considering an intermediate climate change scenario (RCP4.5, MPI-ESM1-2-HR). (**c,d**) The 'risk of non-adaptedness' (RONA) with respect to changes in bio1 and bio12 was modelled using 2,758 and 2,207 putatively adaptive variants, respectively. Colors indicate estimated RONA values. The size of the circles marks the number of individuals per population (n=2-10).

Our genomic offset results demonstrate the potential for employing putatively adaptive genetic variation to predict general adaptive capacity to future climate conditions of a largely autochthonous forest tree species. Notably, differences in the predicted values of future performance of geographically adjacent populations stress the relevance of a high spatial resolution. Our results suggest that each forest stand should be assessed individually rather than relying on interpolated data, as a continuous distribution of adaptive alleles may not always be warranted. Another important consideration concerns variable selection. The future offset of adaptive variants can strongly differ between different environmental variables as exemplified for bio1 and bio12 (Fig. 3c, d). The strength of the genetic association may not always be sufficient for determining the most relevant environmental variables. Under different environmental conditions different environmental variables and thus, different adaptive alleles will be most important for shaping the species range.

Before practical implementation, the empirical validation of the robustness of genomic offset predictions is of critical importance. To this end, the effect of adaptive variants on tree performance can be assessed in common gardens in a space-for-time approach[16]. If the genomic offset measures are representative of the performance of populations under future climate conditions, offset values calculated between the origins of the populations and the common garden should correlate with fitness in the garden. However, the site should ideally feature conditions similar to the conditions predicted under climate change. Our common garden, on the contrary, represents an environment with low levels of abiotic stress. Winter and summer temperatures are moderate and water availability is near optimal throughout the year (Supplementary Table 4). Accordingly, trait differentiation between populations is minimal with the population origin (provenance) explaining less than 3% of the total phenotypic variation of the selected trees in stem circumference (one-way ANOVA, $p=0.0017$, adjusted $R^2=0.026$), a highly integrated trait that can serve as a proxy for fitness and was measured for all resequenced individuals (Supplementary Table 5). Nevertheless, about half of this variance could be explained by our RONA estimates for bio12 in a linear regression model ($p=0.1$, adjusted $R^2=0.016$). Despite not reaching statistical significance at the 5% level, these data indicate that our estimates are robust in principle. Future validation experiments should focus on common gardens with conditions similar to the ones expected under climate change. Under more stressful hotter and drier conditions, populations are expected to exhibit pronounced differences in climate maladaptation. Also, sample size should be increased to more than 10 individuals per population.

In conclusion, using genomics to predict future species performance and ecosystem stability is a rapidly developing field. Especially in long-lived trees it promises great potential for forest management and conservation. Careful variable selection and rigorous validation will be necessary for practical implementation. Nevertheless, our analyses and previous work[33] using genome-wide data across the range of keystone forest tree species already show the value of adding genomics to the prediction of climate change effects. With increasing genomic and phenotypic data, it will be possible to enhance the prediction models, contribute to conserving and restoring forest as one of the most important carbon sinks, and thus mitigate climate change.

Material and methods

Plant material and sampling. For the sample collection of beech genotypes from across the distribution range, we employed one of 38 common garden experimental sites (provenance trials) planted in 1995 and 1998. Specifically, we used the site near Schädtbek (trial code BU1901), where individuals from 100 different populations are growing in a randomized complete block design[34,35]. 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) from above. Tree circumference at breast height was measured for all resequenced trees in December 2022 (Supplementary Table 5).

DNA extraction and Illumina resequencing. Young leaves were sampled in the field in May and June 2021 and placed on ice in collection microtubes (QIAGEN, Hilden, Germany) in a 96-well format. Samples were stored at -20°C until DNA extraction, following a previously described protocol[36]. DNA sample QC, library preparation and sequencing were performed by Novogene Europe. Of the 900 samples, 26 failed sample QC, probably due to low amounts of starting material. For the remaining 874 samples, Illumina sequencing libraries were prepared and an average of 22.98 Gb of 150 bp paired-end data, which corresponds to approximately 42.5x coverage of the beech reference genome[25], were generated on the Illumina Novaseq 6000 platform with an S4 flow cell. Additional details on the sequencing statistics can be found in Supplementary Table 2.

Variant calling. Variant calling was performed using GATK version 4.0.5.1 following the best practices for germline short variant discovery where possible[37]. After read filtering, which consisted of (1) removing reads containing adapters (5' Adapter: 5'-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT-3', 3' Adapter: 5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCACGGATGACTATCTCGTATGCCGTCTTCTGCTTG-3'), (2) removing reads containing more than 10% undetermined bases, and (3) removing reads of low quality, that is reads with a Qscore<= 5 for more than 50% of the bases, the clean data were mapped against the chromosome-level beech reference genome[25] using bwa-mem version 0.7.12 (ref. [38]) with the following parameters: -k 32 -M -R. Duplicate reads were marked with Picard tool's (v2.26.2) 'MarkDuplicates' (http://broadinstitute.github.io/picard/). Using GATK's 'HaplotypeCaller', gVCF files were generated which were combined by GATK's 'CombineGVCFs'. Finally, joint genotyping was performed using GATK's 'GenotypeGVCFs' (ref. [39]).

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

We then extracted the variant sequencing depth values (DP) using vcftools v0.1.15 (ref. [41]) to visualize the DP distributions. To avoid any potentially hemizygous or duplicated sequence variants not resolved in the reference genome, we chose relatively stringent coverage cutoffs based on the mode of the distribution. We only allowed for DP values from 25% below to 50% above the mode of the distribution, which represent values in between haploid and diploid or diploid and tetraploid coverage, respectively. This resulted in cutoffs from 24.9 to 49.8 for SNPs and from 24.8 to 49.7 for indels. Additionally, we filtered out variants with more than 10% missing data and only kept biallelic variants. This resulted in a final dataset of 11.97 million variants for 874 individuals.

Using PLINK v1.9 (ref. [42]) we first checked for general patterns in our dataset using a principle component analysis (--pca). This analysis highlighted nine individuals, all from the Easternmost provenance from Bulgaria, as outliers (Supplementary Fig. 3a). Since the origin of the Bulgarian provenance (provenance 158) is close to the distribution range of the second beech species in Europe, *Fagus orientalis* (https://www.euforgen.org/species/fagus-orientalis), these individuals may not represent *Fagus sylvatica* and were therefore removed from the dataset. Additionally, we excluded one population from Northern Germany that appeared admixed (provenance 32) and one individual from a German population (individual B9) that was closely related to individuals from a population from Slovakia (provenance 135) and may represent a planting error (Supplementary Fig. 3b).

We then filtered for minor allele frequency of 0.01 and pruned variants in complete linkage disequilibrium (LD) in windows of 100 variants (--indep-pairwise 100 25 0.99). Variants with extreme deviation from Hardy-Weinberg equilibrium (p < 1e-08) were removed[43]. Finally, we checked for individuals with excessive levels of missing data or heterozygosity, as previously described[44,45]. We removed a single individual with high levels of missing data (78% vs less than 1.6% for all other individuals). Additionally, we removed eight individuals based on high levels of heterozygosity, exceeding three standard deviations from the mean.

To identify related individuals, we used the program KING v2.3.0 (ref.[46]) with 540 thousand mostly independent variants (LD < 0.2 in windows of 50 variants). This analysis identified 192 individuals with a relatedness of second degree (first cousins) or higher, which we removed using the function '--unrelated --degree 2' Our final dataset comprised 3.68 million variants, of which 540,566 variants exhibit LD values below 0.2 in windows of 50 variants, for 653 "unrelated" individuals.

Genotype-environment association (GEA) and genomic offset analyses. For the analysis of genotype-environment associations we first extracted 19 bioclimatic variables from the WorldClim database[26] (v2.1) with a resolution of 5x5 km for all 98 populations using the R packages geodata[47] and terra[48]. This resolution approximately matches the precision of the geographic coordinates of our populations. Employing the R package LEA v3.10.1 (ref.[49]), we imported our final 540,566 independent variants using the function 'ped2lfmm'. Population genetic structure was evaluated with a principal component analysis and an admixture analysis that are both implemented in LEA and called by the pca()- and snmf()-function, respectively. Both methods indicated 3 genetic clusters in our data and thus, we chose K = 3 latent factors to account for confounding effects caused by population structure. Missing genotype data were imputed using the impute()-function in LEA. We fitted the latent factor mixed model (LFMM) using the lfmm()-function in LEA with 5 repetitions, 10 000 iterations and a burnin of 5 000. The bioclim variables were used as environmental file. P-values were then extracted using the lfmm.pvalues()-function and a conservative Bonferroni significance threshold was applied.

To estimate genomic offset of our studied populations, we employed the 'risk of non-adaptedness' (RONA) measure[30,31] for two bioclimatic variables, i.e. mean annual temperature (bio1) and total annual precipitation (bio12). In order to calculate the population-specific allele frequencies of significant variants, RONA requires the p-value of every variant as determined by the LFMM analysis for the respective bioclim variable and the imputed genotype matrix as used for the LFMM analysis. Further, present bioclim variables as well as a future climate scenario are needed. For every population and for our common garden in Schädtbek (54°3' "N 10°28' "E) we extracted future (2081 – 2100) environmental data for bio1 and bio12 from the WorldClim CMIP6 for the MPI-ESM1-2-HR model[50] with a moderate (RCP 4.5) and a more pessimistic (RCP 8.5) shared socioeconomic pathway (sscp) with a 2.5-minutes resolution (5x5 km). All environmental data are given in Supplementary Table 4.

Statistics and data visualization. For all statistical analyses and data visualization we used R v4.2.2 (ref.[51]) and the ggplot2 package[52].

Data availability

All sequencing data are being uploaded to NCBI's SRA. For access to any of the processed data, please contact the corresponding author to discuss options for data transfer.

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, K. Liepe, M Liesebach and other members of the

Thünen Institute of Forest Genetics for helpful comments and discussions. This work was funded by a core grant from the Thünen Institute.

Author contributions

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

Competing interests statement

The authors declare that they have no competing interests.

References

1. Savolainen, O., Lascoux, M. & Merila, J. Ecological genomics of local adaptation. *Nat Rev Genet* **14**, 807-20 (2013).
2. Capblancq, T., Fitzpatrick, M.C., Bay, R.A., Exposito-Alonso, M. & Keller, S.R. Genomic prediction of (mal)adaptation across current and future climatic landscapes. *Annual Review of Ecology, Evolution, and Systematics* **51**, 245-269 (2020).
3. Erb, K.H. *et al.* Unexpectedly large impact of forest management and grazing on global vegetation biomass. *Nature* **553**, 73-76 (2018).
4. Bar-On, Y.M., Phillips, R. & Milo, R. The biomass distribution on Earth. *Proc Natl Acad Sci U S A* **115**, 6506-6511 (2018).
5. Bastin, J.F. *et al.* The global tree restoration potential. *Science* **365**, 76-79 (2019).
6. Cook-Patton, S.C. *et al.* Mapping carbon accumulation potential from global natural forest regrowth. *Nature* **585**, 545-550 (2020).
7. Dyderski, M.K., Paź, S., Frelich, L.E. & Jagodziński, A.M. How much does climate change threaten European forest tree species distributions? *Global Change Biology* **24**, 1150-1163 (2018).
8. Martinez Del Castillo, E. *et al.* Climate-change-driven growth decline of European beech forests. *Commun Biol* **5**, 163 (2022).
9. Thomas, E. *et al.* Genetic considerations in ecosystem restoration using native tree species. *Forest Ecology and Management* **333**, 66-75 (2014).
10. Exposito-Alonso, M. *et al.* Natural selection on the Arabidopsis thaliana genome in present and future climates. *Nature* **573**, 126-129 (2019).
11. Capblancq, T. *et al.* Climate-associated genetic variation in Fagus sylvatica and potential responses to climate change in the French Alps. *Journal of Evolutionary Biology* **33**, 783-796 (2020).
12. Gougherty, A.V., Keller, S.R. & Fitzpatrick, M.C. Maladaptation, migration and extirpation fuel climate change risk in a forest tree species. *Nature Climate Change* **11**, 166-171 (2021).
13. Waldvogel, A.-M. *et al.* Evolutionary genomics can improve prediction of species' responses to climate change. *Evolution Letters* **4**, 4-18 (2020).
14. Aguirre-Liguori, J.A., Ramírez-Barahona, S. & Gaut, B.S. The evolutionary genomics of species' responses to climate change. *Nature Ecology & Evolution* **5**, 1350-1360 (2021).
15. Exposito-Alonso, M. Understanding local plant extinctions before it's too late: bridging evolutionary genomics with global ecology. *New Phytol* (2023).
16. Fitzpatrick, M.C., Chhatre, V.E., Soolanayakanahally, R.Y. & Keller, S.R. Experimental support for genomic prediction of climate maladaptation using the machine learning approach Gradient Forests. *Mol Ecol Resour* **21**, 2749-2765 (2021).
17. Gain, C. *et al.* A quantitative theory for genomic offset statistics. *bioRxiv*, 2023.01.02.522469 (2023).
18. Bay, R.A. *et al.* Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science* **359**, 83-86 (2018).

19. Rhone, B. *et al.* Pearl millet genomic vulnerability to climate change in West Africa highlights the need for regional collaboration. *Nat Commun* **11**, 5274 (2020).
20. Hickler, T. *et al.* Projecting the future distribution of European potential natural vegetation zones with a generalized, tree species-based dynamic vegetation model. *Global Ecology and Biogeography* **21**, 50-63 (2012).
21. Eisenstein, M. Illumina faces short-read rivals. *Nat Biotechnol* **41**, 3-5 (2023).
22. Hoffmann, A.A., Weeks, A.R. & Sgro, C.M. Opportunities and challenges in assessing climate change vulnerability through genomics. *Cell* **184**, 1420-1425 (2021).
23. Rellstab, C., Dauphin, B. & Exposito-Alonso, M. Prospects and limitations of genomic offset in conservation management. *Evol Appl* **14**, 1202-1212 (2021).
24. Laruson, A.J., Fitzpatrick, M.C., Keller, S.R., Haller, B.C. & Lotterhos, K.E. Seeing the forest for the trees: Assessing genetic offset predictions from gradient forest. *Evol Appl* **15**, 403-416 (2022).
25. Mishra, B. *et al.* A chromosome-level genome assembly of the European beech (Fagus sylvatica) reveals anomalies for organelle DNA integration, repeat content and distribution of SNPs. *Front Genet* **12**, 691058 (2021).
26. Fick, S.E. & Hijmans, R.J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**, 4302-4315 (2017).
27. Dauphin, B. *et al.* Re-thinking the environment in landscape genomics. *Trends in Ecology & Evolution* **38**, 261-274 (2023).
28. Frichot, E., Schoville, S.D., Bouchard, G. & Francois, O. Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol* **30**, 1687-99 (2013).
29. Barton, N.H. The "New Synthesis". *Proc Natl Acad Sci U S A* **119**, e2122147119 (2022).
30. 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).
31. 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).
32. Dauphin, B. *et al.* Genomic vulnerability to rapid climate warming in a tree species with a long generation time. *Global Change Biology* **27**, 1181-1195 (2021).
33. 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).
34. 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).
35. Liesebach, M. International beech provenance trial 1993/95 - site Schädtbek (Bu19-1). *Thünen Report* **62**, 131-138 (2017).
36. 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).
37. Poplin, R. *et al.* Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* (2018).
38. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
39. van der Auwera, G. & O'Connor, B.D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*, (O'Reilly Media, Incorporated, 2020).
40. 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).
41. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-8 (2011).
42. 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).
43. 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).

44. Anderson, C.A. *et al.* Data quality control in genetic case-control association studies. *Nat Protoc* **5**, 1564-73 (2010).
45. 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).
46. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-2873 (2010).
47. Hijmans, R.J., Barbosa, M., Ghosh, A. & Mandel, A. geodata: Download Geographic Data. *R package* (2023).
48. Hijmans, R.J., Bivand, R., Pebesma, E. & Sumner, M.D. terra: Spatial Data Analysis. *R package* (2023).
49. Gain, C. & François, O. LEA 3: Factor models in population genetics and ecological genomics with R. *Mol Ecol Resour* **21**, 2738-2748 (2021).
50. 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).
51. R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* (2022).
52. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*, (Springer-Verlag New York, 2016).