Received: 5 February 2024 | Revised: 19 April 2024 | Accepted: 25 April 2024

DOI: 10.1111/ppl.14334

ORIGINAL RESEARCH

Physiologia Plantarum

A genome-wide genetic association study reveals SNPs significantly associated with environmental variables and specific leaf area in European beech

Markus Müller[1,2] | Christoph Leuschner[3,6] | Greta Weithmann[3] | Robert Weigel[3,7] | Bat-Enerel Banzragch[3,4] | Wilfried Steiner[5] | Oliver Gailing[1,2,6]

[1]University of Göttingen, Forest Genetics and Forest Tree Breeding, Göttingen, Germany

[2]Center for Integrated Breeding Research (CiBreed), University of Goettingen, Göttingen, Germany

[3]Department Plant Ecology and Ecosystems Research, University of Göttingen, Göttingen, Germany

[4]Applied Vegetation Ecology, Faculty of Environment and Natural Resources, University of Freiburg, Freiburg, Germany

[5]Department Forest Genetic Resources, Northwest German Forest Research Institute, Hann. Münden, Germany

[6]Center of Sustainable Land Use (CBL), Georg-August-University Göttingen, Göttingen, Germany

[7]Ecological-Botanical Garden, University of Bayreuth, Bayreuth, Germany

Correspondence
Oliver Gailing,
Email: oliver.gailing@uni-goettingen.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: 433288081; Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz, Grant/Award Numbers: 2218WK34A4, 22WC415001; Bundesministerium für Ernährung und Landwirtschaft, Grant/Award Numbers: 2218WK34A4, 22WC415001

Edited by J.-F. Mao

Abstract

European beech is negatively affected by climate change and a further growth decline is predicted for large parts of its distribution range. Despite the importance of this species, little is known about its genetic adaptation and especially the genetic basis of its physiological traits. Here, we used genotyping by sequencing to identify SNPs in 43 German European beech populations growing under different environmental conditions. In total, 28 of these populations were located along a precipitation and temperature gradient in northern Germany, and single tree-based hydraulic and morphological traits were available. We obtained a set of 13,493 high-quality SNPs that were used for environmental and SNP-trait association analysis. In total, 22 SNPs were identified that were significantly associated with environmental variables or specific leaf area (SLA). Several SNPs were located in genes related to stress response. The majority of the significant SNPs were located in non-coding (intergenic and intronic) regions. These may be in linkage disequilibrium with the causative coding or regulatory regions. Our study gives insights into the genetic basis of abiotic adaptation in European beech, and provides genetic resources that can be used in future studies on this species. Besides clear patterns of local adaptation to environmental conditions of the investigated populations, the analyzed morphological and hydraulic traits explained most of the explainable genetic variation. Thus, they could successfully be altered in tree breeding programs, which may help to increase the adaptation of European beech to changing environmental conditions in the future.

1 | INTRODUCTION

During their long lifespan, forest trees encounter changing and stressful environmental conditions. Genetic variation in relevant genes is essential for their long-term evolutionary adaptation in the face of climate change. European beech (*Fagus sylvatica* L.) is an economically and ecologically important deciduous tree species in Central Europe. It is regarded as moderately drought-sensitive and more vulnerable than most of the native deciduous tree species of the genera *Quercus*, *Carpinus*, *Tilia*, *Sorbus* and *Fraxinus* (Leuschner, 2020). European beech is negatively affected by climate change in large regions from the south to the center of its distribution range mostly due to a deterioration of the climatic water balance in summer (Knutzen et al., 2017; Martinez del Castillo et al., 2022; Weigel et al., 2023). Martinez del Castillo et al. (2022) found a growth decline of European beech in a wide range of its distribution area from 1955 to 2016 and a significant further growth decline is predicted for the future. In some parts of Germany, an increasing dieback of mostly older trees has been observed since the extreme hot drought in 2018 (Neycken et al., 2022; NW-FVA, 2019). Therefore, a better understanding of the genetic basis of adaptative traits and adaptation to changing environmental conditions, in particular increasing summer drought exposition, is necessary.

Most studies that investigated abiotic genetic adaptation in European beech used the candidate gene approach to identify Single Nucleotide Polymorphisms (SNPs) that are significantly associated with environmental variables and/or adaptive traits. For instance, SNPs located in candidate genes were used to conduct gene–environment association (GEA) analyses in European beech populations growing along environmental gradients (Csilléry et al., 2014; Cuervo-Alarcon et al., 2018; Krajmerová et al., 2017; Postolache et al., 2021). Other studies used SNPs located in candidate genes to detect associations between SNPs and adaptive traits (Cuervo-Alarcon et al., 2021; Krajmerová et al., 2017; Müller et al., 2017). There are only few studies that used genome-wide SNP sets to analyze environmental adaptation in European beech (e.g., Capblancq et al., 2020; Meger et al., 2021; Pfenninger et al., 2021). Often, only a small amount of phenotypic variation can be explained by the identified SNPs. This is likely due to the polygenic nature of most adaptive traits in forest tree species, in which many genes with small effects are responsible for trait expression (Aitken et al., 2008; Le Corre & Kremer, 2012; Lind et al., 2018). Therefore, different methods were developed that take joint marker effects and also multivariate environmental data into account (Capblancq & Forester, 2021). For instance, Capblancq et al. (2020) used a redundancy analysis (RDA) to perform an association analysis between a genome-wide set of SNPs and a set of climatic variables of beech populations in the French Alps. European beech contains high neutral and potentially adaptive intra-population genetic variation, which is a good basis for adaptation to changing environmental conditions (Meger et al., 2021; Müller et al., 2018). Genetic association studies revealed SNPs that are associated with climatic variables (Cuervo-Alarcon et al., 2018; Meger et al., 2021). Also, SNPs significantly associated with traits such as bud burst timing, chlorophyll fluorescence parameters, or drought damage status were identified (Cuervo-Alarcon et al., 2021; Krajmerová et al., 2017; Müller et al., 2015; Pfenninger et al., 2021). However, less is known about the genetic basis of tree hydraulic and leaf morphological traits. Since studies detected differences in tree hydraulic/morphological traits (e.g., stomatal conductance, root/shoot ratio, shoot water potential) among provenances originating from regions with different environmental conditions (García-Plazaola & Becerril, 2000; Leuschner, 2020; Peuke et al., 2002; Rose et al., 2009), local adaptation, and hence, a genetic control of these traits can be assumed. Our knowledge of differences in tree physiological traits is mainly based on seedling experiments and more research on adult trees is needed (Leuschner, 2020). To our knowledge, there are so far no genome-wide association studies (GWAS) that have identified genetic variation associated with tree hydraulic/morphological traits in European beech.

This study aims to provide a genome-wide set of SNP markers to identify genes and SNPs that are involved in abiotic adaptation and associated with tree hydraulic/morphological traits in European beech and to evaluate their potential to increase the adaptation of European beech to changing environmental conditions in breeding programs. We used genotyping by sequencing to identify genome-wide SNPs in 43 adult beech populations from different regions in Germany, which differ in their environmental conditions. Twenty-eight of these populations were located along a precipitation and temperature gradient in northern Germany. For these, single tree-based data for various hydraulic and morphological variables were available from previous studies. We used these variables to conduct a genome-wide association study. Further, we conducted a gene–environment association (GEA) analysis using publicly available climate and soil data of all 43 populations to test for the potential of local genetic adaptation to different abiotic habitat conditions. For our analyses, we used both univariate and multivariate approaches to identify SNPs that might be involved in adaptation in European beech.

2 | MATERIALS AND METHODS

2.1 | Plant material

The study was carried out in 43 European beech (*Fagus sylvatica* L.) populations (forest stands) from all over Germany (Table 1, Figure 1). The populations were selected to cover different climatic conditions. Twenty-eight populations were located along a precipitation (522–886 mm mean annual precipitation (MAP), 1991–2018 average) and temperature (9.0–10.0°C mean annual temperature (MAT)) gradient in northern Germany as described in Weithmann et al. (2022a). For these populations, individual-based tree hydraulic and morphological variables were available (see below). The remaining 15 populations were part of the project "Biodiversity Exploratories" (https://www.biodiversity-exploratories.de/en/) and are located in three widely separated areas in Germany (northeast, central and southwest Germany) that differ in climatic conditions. At each of the three sites, five

TABLE 1 Overview of analyzed populations. "Gradient" refers to populations from the temperature/precipitation gradient in northern Germany for which tree physiological traits were available. "Exploratories" refers to populations from the Biodiversity Exploratories project. EQ: Ellenberg's climate quotient.

ID	Location set	latitude	longitude	EQ
B2	Gradient	54.4287	9.653	19
B3	Gradient	54.1547	10.5594	23
B4	Gradient	51.4997	11.9165	37
B5	Gradient	51.7496	12.2095	36
B6	Gradient	51.6811	12.5857	32
B7	Gradient	52.0416	12.3792	32
B9	Gradient	52.4461	13.0745	32
B10	Gradient	53.4128	7.7775	21
B11	Gradient	53.5998	8.7668	22
B12	Gradient	53.7829	8.6215	21
B13	Gradient	53.853	8.612	21
B14	Gradient	52.8964	13.8539	33
B15	Gradient	53.1133	12.8536	30
B16	Gradient	53.1996	12.7374	30
B17	Gradient	52.696	13.4046	31
B18	Gradient	53.6747	11.7339	28
B19	Gradient	53.715	14.1382	33
B20	Gradient	52.6207	11.2298	29
B21	Gradient	52.4039	11.2606	32
B22	Gradient	53.1226	10.8196	27
B23	Gradient	53.1722	9.9528	22
B24	Gradient	52.8312	10.317	24
B25	Gradient	54.4167	13.5552	32
B26	Gradient	54.0201	12.4801	31
B27	Gradient	53.4612	9.9075	22
B28	Gradient	53.6233	9.7681	22
B29	Gradient	54.0554	10.7229	24
B30	Gradient	53.9619	10.1255	22
SEW47	Exploratories	53.07	13.77	31
SEW45	Exploratories	53.05	13.84	32
SEW48	Exploratories	53.05	13.84	32
SEW36	Exploratories	52.95	13.75	31
SEW49	Exploratories	52.89	13.89	32
AEW35	Exploratories	48.45	9.42	19
AEW17	Exploratories	48.4	9.24	16
AEW49	Exploratories	48.45	9.48	18
AEW21	Exploratories	48.48	9.32	18
AEW41	Exploratories	48.36	9.4	18
HEW19	Exploratories	51.2	10.34	20
HEW34	Exploratories	51.08	10.45	21
HEW23	Exploratories	51.28	10.21	18
HEW44	Exploratories	51.33	10.37	18
HEW27	Exploratories	51.36	10.52	17

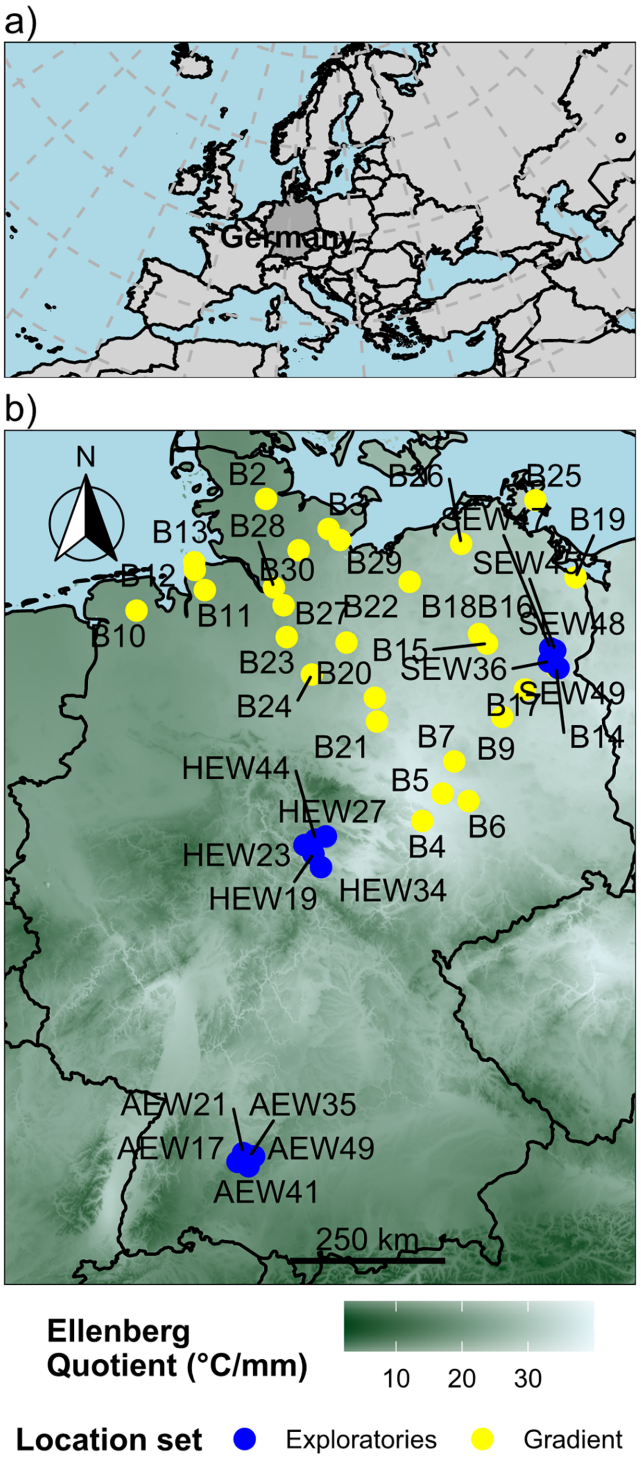

FIGURE 1 Overview of the sampling area. (a) Displayed is the location of Germany within Europe and (b) the location of the studied populations within Germany plotted over Ellenberg's climate quotient. "Gradient" refers to populations from the temperature/precipitation gradient in northern Germany for which tree physiological traits were available. "Exploratories" refers to populations from the Biodiversity Exploratories project.

populations were sampled. Leaves of 10 individuals per population (9 in population B16) were sampled (in total, 429 individuals). Leaves were kept frozen at −20°C until DNA extraction.

2.2 | Measurement of environmental and tree hydraulic/morphological variables

Environmental variables for the 43 European beech populations were downloaded from the WorldClim (Fick & Hijmans, 2017) gridded climate dataset (1970–2000 average) at a resolution of 30 arc seconds (~1km^2) and SoilGrids 2.0 (Poggio et al., 2021) (250 m raster resolution) databases using the raster v.3.5–15 R package (Hijmans, 2022) and soilDB v.2.6.13 R package (Beaudette et al., 2022), respectively. Based on the WorldClim variables, additional composite variables were calculated: mean spring (April to June) precipitation and temperature, mean growing season (April to September) precipitation and temperature, as well as Ellenberg's climate quotient, which relates summer heat and moisture conditions (1000 * (July temperature / MAP)) (Ellenberg, 1963). Individual tree hydraulic and morphological traits (e.g., vessel density, specific leaf area, δ13C) for the 28 populations growing along the precipitation and temperature gradient were obtained from the BEECHLIMITS project (Weigel et al., 2023; Weithmann et al., 2022a; Weithmann et al., 2022b; Weithmann et al., 2022c). The drought sensitivity of each tree was assessed by correlating annual tree-ring increment with the drought index SPEI (Standardized Precipitation – Evapotranspiration Index; Vicente-Serrano et al., 2010; Weigel et al., 2023). In total, 17 traits were used (Table S1).

2.3 | DNA extraction, genotyping by sequencing and SNP identification

DNA was extracted from leaves using the DNeasy 96 Plant Kit (Qiagen) and sent to LGC Genomics for normalized genotyping by sequencing (nGBS; Arvidsson et al., 2016) and SNP identification. Paired-end sequencing (2 × 150 bp) was conducted on an Illumina NextSeq 550 system with an average number of ca. 1.5 million read pairs per sample. Raw sequencing reads were deposited in the NCBI Sequence Read Archive (SRA) under BioProject number PRJNA1043878. After demultiplexing and quality trimming, the reads were aligned against the *Fagus sylvatica* reference genome v.1.2 (Mishra et al., 2018) using BWA-MEM v.0.7.12. SNP discovery was conducted using Freebayes v.1.2.0 (Garrison & Marth, 2012). An initial filtering of the SNPs comprised the following settings: total number of fully covered SNPs in 10% of samples, MAF ≥0.05, and a minimum read count of 8. The resulting VCF file was used for downstream analyses. The R package vcfR v.1.12.0 (Knaus & Grünwald, 2017) was used to convert the VCF file into the genlight format, and the dartR v.2.3.3 package (Gruber et al., 2018) was used for further filtering of the SNPs (call rate of 0.8, linkage disequilibrium (LD): $R^2 < 0.5$, no monomorphic loci). This final set of 13,493 SNPs was used for the analyses described in the following paragraphs. The SNPs were annotated with SnpEff v.5.1 (Cingolani et al., 2012a). Where necessary, the annotated SNP file was further processed (e.g., filtering for intergenic SNPs or significant SNPs from the association analyses (see below)) with SnpSift v.5.1 (Cingolani et al., 2012b) or re-Searcher v.1.0 (Karabayev et al., 2021). The function of significant SNPs that were identified by the different association analyses (see below) was determined with tbg-tools v.0.2 (Schönnenbeck et al., 2021).

2.4 | Population structure

The STRUCTURE v.2.3.4 software (Pritchard et al., 2000) was used to infer population structure. We conducted different STRUCTURE analyses based on different population and/or marker subsets to identify the overall population structure of the populations and to infer neutral population structure for inclusion in the genetic association analyses (see below). The different settings were 1) all populations with the total filtered SNP set, 2) all populations without potentially planted ones (based on the first STRUCTURE result) with the total SNP set, 3) all populations despite the potentially planted ones with 1088 intergenic SNPs, and 4) the 28 populations from the northern German precipitation and temperature gradient with the 1088 identified putatively neutral intergenic SNPs. In the STRUCTURE software, we used the admixture model, correlated allele frequencies, and a burn-in period of 10,000 and Markov Chain Monte Carlo (MCMC) replicates of 100,000. We tested from *K*1 to *K*5 potential clusters using 5 iterations. We used StrAuto v1.0 (Chhatre & Emerson, 2017) to run STRUCTURE on the high performance computing system of the Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG). STRUCTURE HARVESTER v.0.6.94 (Earl & vonHoldt, 2012) was used to infer the most likely number of clusters (*K*) based on the Δ*K* method (Evanno et al., 2005). Finally, CLUMPAK (Kopelman et al., 2015) was applied for summation and graphical representation of the STRUCTURE results. Further, an analysis of molecular variance (AMOVA) over all populations based on 1000 permutations was conducted with the poppr v.2.9.4 R package (Kamvar et al. 2014; Kamvar et al. 2015), and a pairwise F_{ST} matrix based on 1000 bootstraps among loci was calculated using the dartR v.2.3.3 R package (Gruber et al., 2018).

2.5 | Environmental association analysis

The STRUCTURE analysis revealed an exceptionally high differentiation for five of the populations, which indicates that the populations were planted from introduced seed material in the past. Since it cannot be expected that these populations are adapted to the local environmental conditions, they were excluded from the environmental association analysis (populations B2, B5, B7, B9, and B13). A principle component analysis (PCA) was conducted with the environmental variables (climate and soil, see above) using the prcomp R function (R Core Team, 2021), and the first 4 principal components (PCs, hereafter called environmental PCs) were used for the environmental association analyses. Correlation (Spearman's rank correlation coefficient) between the environmental PCs and the environmental variables was calculated and tested for significance ($p < 0.05$ after correction for multiple testing (false discovery rate, fdr) with the corr.test function of the psych v2.1.9 R package (Revelle, 2021) in order to identify environmental variables that are highly correlated with the PCs.

Three different approaches were used for the environmental association analysis: Bayenv2.0 (Günther & Coop, 2013), latent factor mixed model (lfmm) (Frichot et al., 2013) implemented in the LEA v.3.2.0 R package (Frichot & François, 2015), and the multivariate approach partial redundancy analysis (pRDA) implemented in the vegan v.2.5–7 R package (Oksanen et al., 2020). Bayenv2.0 calculates standardized allele frequencies and tests for correlations between those frequencies and environmental variables while accounting for spurious correlations due to population history and gene flow (Günther & Coop, 2013). For the latter, a covariance matrix based on potentially neutral genetic markers is included in the model. We used the 1088 intergenic SNPs to calculate this matrix based on 100,000 iterations and 10 independent runs, which were summarized into a mean covariance matrix. Bayenv2.0 was finally run using 100,000 iterations. SNPs with Bayes factor ≥3 and included in the top 1% of Pearson correlation coefficients were considered as significantly associated with the environmental PCs.

Lfmm does not allow for missing data. Therefore, we imputed missing data using the impute function implemented in the LEA R package (Frichot & François, 2015). For the association analysis, we ran the lfmm function with 50,000 burn-in cycles and 100,000 iterations and 10 repetitions. The number of latent factors *K* was set to 2 based on the STRUCTURE results (see above) and the results of the snmf function implemented in the LEA R package, which estimates individual ancestry coefficients. The lfmm.pvalues function was used to summarize the *z-scores* of the different runs and to calculate p-values. The p.adjust R function (R Core Team, 2021) was used to adjust the p-values for multiple testing based on the Benjamini and Hochberg method (fdr) (Benjamini & Hochberg, 1995). SNPs with adjusted p-values of $p \leq 0.05$ were considered as significant.

For the pRDA analysis, we followed the method described in Capblancq et al. (2018) and Capblancq and Forester (2021), in which outlier loci are identified based on their extremeness along a distribution of Mahalanobis distances (Figure S1). Population structure was included in the model as using individual *q*-values obtained from the STRUCTURE analysis based on intergenic SNPs (see above).

2.6 | SNP – trait association analysis

Correlations between trait variables were tested using the corr.test function of the psych v.2.1.9 R package (Revelle, 2021). For all pairs of traits that were highly correlated (Spearman's rank correlation coefficient ≥0.7) (Table S2), one of the traits was excluded from the analysis. In total, three variables ("theoretical xylem specific conductivity", "mean vessel diameter", and "vessel grouping index") were excluded but instead, "lumen-to-sapwood area ratio", "hydraulically-weighted vessel diameter", and "vessel solitary faction" were retained. Further, nine individuals with missing data for some of the variables were excluded. Our analysis focused on natural beech populations and not populations growing under controlled environments such as in common garden experiments. Thus, the relative contribution of environment and traits to the observed genetic variation could not directly be inferred (in common garden experiments, the observed differences in trait expression are largely due to genetic differences since the plants are growing under the same environmental conditions). Therefore, we used variance partitioning with pRDA to identify the contribution of the different factors (neutral genetic population structure, geography/climate, and tree hydraulic/morphological traits) to the total genetic variation (Capblancq & Forester, 2021). Neutral genetic population structure was included in the model as *q*-values from STRUCTURE based on intergenic SNPs, and geography was included as longitudinal and latitudinal coordinates of the individuals. Geography was correlated with climatic variables, and hence, it can also be seen as a proxy for the different climatic conditions under which the individuals are growing. In order to separate the effects of geography, genetic population structure, and traits, we ran four different models (Table 2): (1) the full model including all variables, (2) a model for "pure" traits including the tree hydraulic/morphological variables while controlling (via the condition() argument in the rda function) for geography and population structure, (3) a "pure structure" model including population structure while controlling for traits and geography, and (4) a "pure geography" model including geography while controlling for traits and population structure. We used TASSEL v.5 (Bradbury et al., 2007) and GAPIT v.3 (Wang & Zhang, 2021) to identify SNPs that are significantly associated with traits. In both programs, we used a general linear model (GLM) and a mixed linear model (MLM). In the GLM, *q*-values from STRUCTURE and geography (see above) were included as covariates to account for population structure and climatic/geographic effects, respectively. In the MLM, additionally a kinship matrix was included to account for relationships among individuals (Figure S2). The kinship matrix was calculated with TASSEL and visualized with the ASRgenomics v.1.1.4 R package (Gezan et al., 2022). GAPIT provides p-values adjusted for multiple testing. Nominal p-values revealed by TASSEL were adjusted using the p.adjust R function as described above. FDR-corrected p-values ≤ 0.05 were used to define significant SNP-trait associations.

3 | RESULTS

3.1 | Genotyping by sequencing and SNP identification

Sequencing revealed ca. 1.5 million reads per sample and the mapping rate to the beech genome was 95.7%. In total, 463,352 SNPs were discovered. After the filtering steps described above, a final set of 13,493 SNPs was obtained. A file containing these SNPs can be found in Suppl. Material (data file S1).

3.2 | Population structure

The STRUCTURE analysis based on the total filtered SNP set revealed $K = 3$ as the most likely number of clusters. Only weak population structure was detected (Figure 2A), whereas the populations from the southern parts of Germany (AEW) were most differentiated. Running STRUCTURE without the potentially planted populations (five populations, see above)

TABLE 2 Results of variance partitioning with pRDA. For each model the p value, the percentage of explainable variance, and the percentage of total variance is shown.

Partial RDA models	Variance	p	Percentage (%) of explainable variance	Percentage (%) of total variance
Full model: F ~ traits + geo. + struct.	344.7	0.001	100.00	8.36
Pure traits: F ~ traits| (geog. + struct.)	246	0.001	71.37	5.97
Pure structure: F ~ struct. | (traits + geog.)	40	0.001	11.60	0.97
Pure geography: F ~ geog. | (traits + struct.)	43	0.001	12.47	1.04
Confounded traits/structure/geography	15.7		4.56	0.38
Total unexplained	3778.2			91.64
Total variance	4122.9			100.00

F: genetic variation, traits: tree hydraulic/morphological traits, geo.: geography, struct.: neutral genetic population structure.

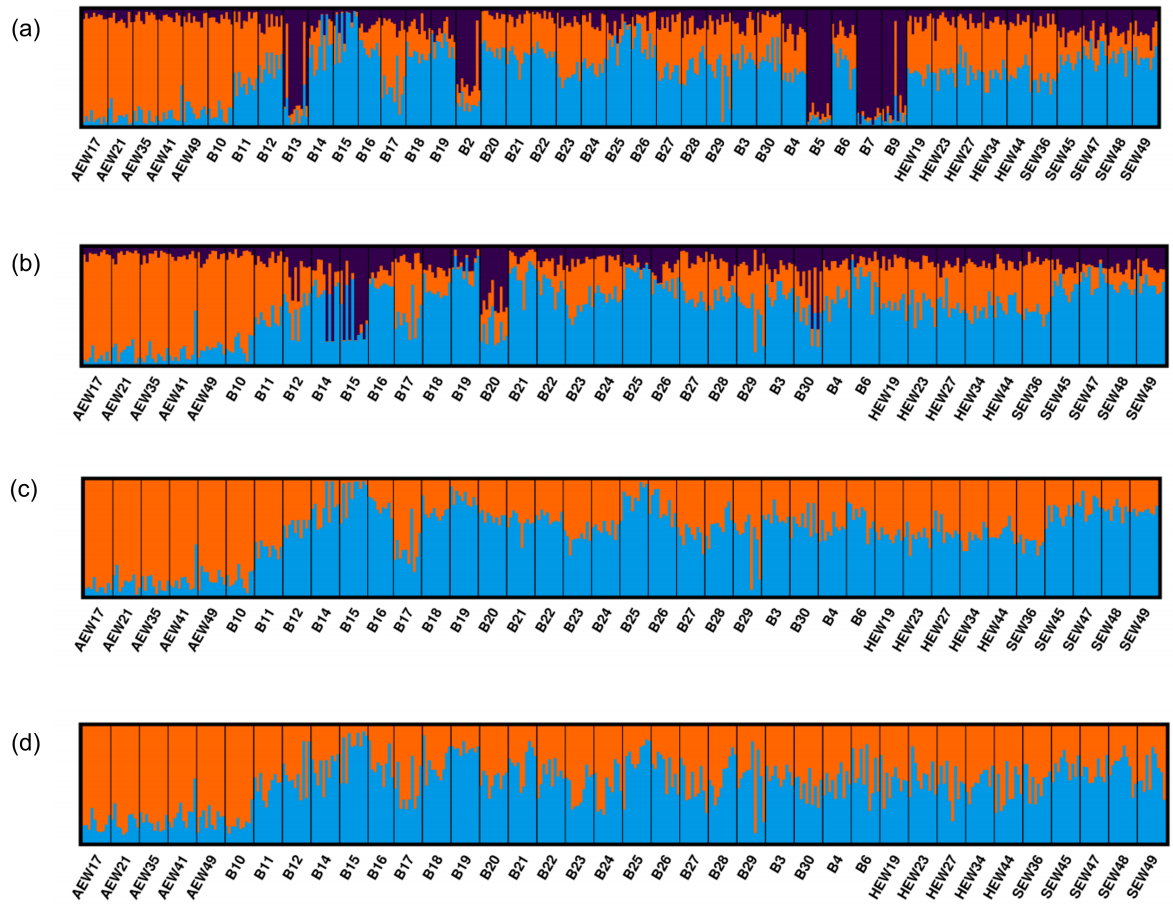

FIGURE 2 Population structure of the analyzed populations. The clusters (*K*) are represented by different colors. (a) Population structure based on the total filtered SNP set for $K = 3$, population structure based on the total filtered SNP set after exclusion of potentially planted populations for $K = 3$ (b) and $K = 2$ (c). Population structure based on intergenic SNPs for $K = 2$ (d).

also revealed $K = 3$ as the most likely number of clusters (Figure 2B). The snmf function implemented in the LEA R package, however, revealed $K = 2$ as the most likely number of clusters (Figure 2C). $K = 2$ was also the most likely number for the other two STRUCTURE analyses based on intergenic SNPs (Figure 2D). The AMOVA revealed that 2.7% of the variation was found among populations, 8.5% between samples within populations, and 88.8% within samples (Table S3). Mean F_{ST} among populations was 0.027 (data file S2).

3.3 | Environmental association analysis

The first 4 environmental PCs together explained 94.4% of the variance in the environmental data. PC1 was negatively correlated with several precipitation variables and positively correlated with several temperature variables (Table S4). For PC2 the strongest positive correlations were found with temperature seasonality and annual temperature range, and the strongest negative correlations with minimum

TABLE 3 Overview of SNPs significantly associated with specific leaf area (SLA) or environmental PCs.

SNP	Trait	Chromosome	Class	Annotation
scaffold119_size409709_63521	PC1	7	Intronic	vesicle-fusing_ATPase-like
scaffold27_size635005_579987	PC1	4	Intergenic	intergenic_region
scaffold468_size229126_58274	PC1	1	Intergenic	intergenic_region
scaffold349_size263893_101889	PC2	7	Intergenic	intergenic_region
scaffold733_size243116_78220	PC2	3	Intronic	serine_hydroxymethyltransferase_6
scaffold153_size597108_249560	PC3	6	Intronic	protein_NYNRIN-like
scaffold1142_size245227_149635	PC4	3	synonymous	GDP-mannose_transporter_GONST3-like
scaffold1327_size114253_36968	PC4	3	Intronic	uncharacterized_mitochondrial_g00810-like
scaffold161_size363313_240927	PC4	5	synonymous	UDP-glucuronate_4-epimerase_6
scaffold285_size289019_189215	PC4	5	Intronic	electron_transfer_flavo_-ubiquinone_mitochondrial_isoform_X1
scaffold1065_size135957_37986	SLA	3	synonymous	uncharacterized protein LOC109000868
scaffold1121_size131042_17041	SLA	1	Intronic	rab escort 1
scaffold1271_size118169_110362	SLA	6	Intronic	eukaryotic initiation factor 4A-8
scaffold1326_size114272_71215	SLA	3	Intronic	uncharacterized abhydrolase domain-containing DDB_G0269086-like
scaffold1785_size128954_116982	SLA	6	Intronic	ubiquitin carboxyl-terminal hydrolase 22-like
scaffold2660_size57589_42688	SLA	3	Intergenic	intergenic_region
scaffold508_size218371_156893	SLA	1	Intergenic	intergenic_region
scaffold912_size271288_190268	SLA	10	synonymous	synaptotagmin-3-like isoform X1
scaffold116_size412086_359247	SLA	11	Intergenic	intergenic_region; 248 bp distance to the next gene (Casein kinase II subunit alpha-2)
scaffold1441_size107806_32277	SLA	5	Intergenic	intergenic_region
scaffold812_size164009_151255	SLA	5	Intergenic	intergenic_region
scaffold719_size176289_97941	SLA	3	Intergenic	intergenic_region

The SNP position (bp) on the scaffold is indicated as the number after the second underscore of the SNP ID.

and maximum temperatures. PC3 was negatively correlated with mean and maximum temperatures in February/March. Finally, PC4 was negatively correlated with precipitation seasonality, cation exchange capacity of the soil, and soil pH. Based on the three different methods used for the environmental association analysis, 300 different SNPs were found to be associated with at least one of the three environmental PCs. The three methods revealed different numbers of significant SNPs (Figure S3). The SNPs that were found to be significantly associated with a given environmental PC with all three methods were regarded as the most reliable associations. These were 3 SNPs for PC1, 2 SNPs for PC2, 1 SNP for PC3, and 4 SNPs for PC4 (Figure S3, Table 3). Of these, 5 were intronic SNPs, 3 intergenic SNPs, and 2 synonymous SNPs (Table 3).

3.4 | SNP – trait association analysis

The association analysis based on TASSEL revealed 12 significant SNPs using the GLM and one SNPs using the MLM, while GAPIT detected no significant SNPs for both applied models. All significant SNPs were associated with specific leaf area (SLA) (Table 3). Of these 12 SNPs, 4 were intronic, 6 were intergenic, and 2 were synonymous (Table 3). The SNPs explained between 6.4% (SNP scaffold719_size176289_97941) and 13.4% (SNP scaffold1065_size135957_37986) of the phenotypic variation (mean R^2 of the significant SNPs: 7.8%). We used variance partitioning with pRDA to identify the contribution of the different factors to the total genetic variation. The full model explained 8.36% of the total genetic variance (Table 2). The effect of the "pure" traits while controlling for population structure and geography was significant and explained 5.97% of the total genetic variation and the majority of the variation (71.37%) explained by the full model (Table 2). Population structure and geography alone explained 0.97% and 1.04% of the total genetic variation. The variance that could not be distinguished among the tested parameters in the model (confounded variance) was 0.38%.

4 | DISCUSSION

We used genotyping by sequencing to identify genome-wide SNPs in European beech populations in Germany. After filtering, we obtained a set of 13,493 SNPs for further analyses.

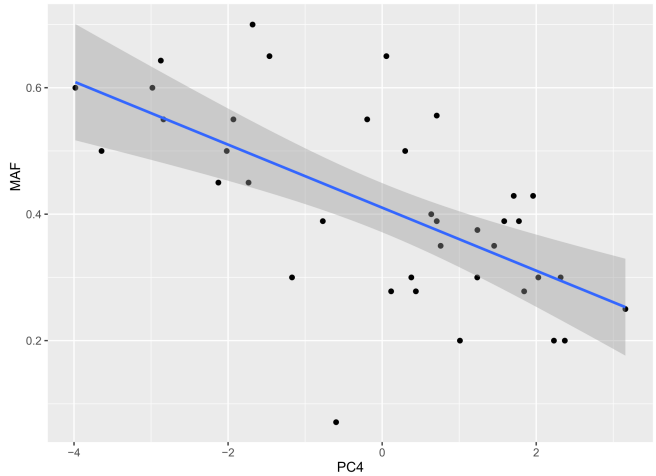

FIGURE 3 Scatterplot between PC4 and the minor allele frequency (MAF) of SNP "scaffold285_size289019_189215". The correlation was significant (ϱ: −0.65, p < 0.0001).

A low population structure was detected among the beech populations, but there are indications of planting (artificial regeneration) for some of them. We identified SNPs that were significantly associated with environmental and soil variables as well as specific leaf area (SLA) (Table 3; Figures 3 and S4). Together, this study gives insights into the genetic basis of adaptation in European beech and the provided SNP set and candidate genes can be used in future studies.

4.1 | Genotyping by sequencing and SNP identification

nGBS was conducted by aiming at 1.5 million reads per sample. In a pilot study, we used six of the individuals from this study to compare sequencing results based on 1.5 million and 3 million reads per sample (data not shown). This pilot study revealed that 1.5 million reads per sample are sufficient and a sequencing based on 3 million reads per sample would not lead to a higher mapping rate or replicate concordance, and would provide only a slightly higher number of SNPs. The final set of 13,493 SNPs obtained after filtering can directly be used in future studies on the genetic adaptation of European beech.

4.2 | Population structure

The complete filtered SNP set was used to infer population structure. The STRUCTURE analysis revealed $K = 3$ as the most likely number of clusters and only weak population structure was detected. Five populations (B2, B5, B7, B9, and B13) showed an unexpected high differentiation and might have been planted in the past. After running the STRUCTURE analysis without these populations, the most likely number of clusters was still $K = 3$ based on the ΔK method (Evanno et al., 2005). The snmf function implemented in the LEA R package (Frichot & François, 2015), however, revealed $K = 2$ as the most likely number of clusters. The highest differentiation (except the potentially planted populations) was observed for the AEW populations from southern Germany. A similar pattern (weak population structure and highest differentiation of these southern populations) was also identified in a previous study on European beech in Germany based on simple sequence repeat (SSR) markers and a small SNP set from candidate genes (Müller et al., 2018). In general, low population structure was often detected in Central European beech populations (Lalagüe et al., 2014; Meger et al., 2021; Müller et al., 2018; Pfenninger et al., 2023; Pluess & Weber, 2012).

4.3 | Environmental and SNP – trait association analyses

In total, 10 SNPs were significantly associated with at least one of the environmental PCs, whereby the SNPs were specific for the respective PC. We conducted a correlation analysis between the environmental PCs and the climatic and soil variables to identify the variables showing the strongest correlation. This analysis revealed that PC1 was strongly correlated with several precipitation and temperature variables. Thus, the three SNPs that are significantly associated with PC1 might be involved in general climatic adaptation. PC2 and PC3 were strongly correlated with variables such as minimum and maximum temperatures or temperature seasonality, and hence, SNPs showing significant associations with these variables might be involved in adaptation to variable and more extreme temperatures. Finally, PC4 was strongest correlated with precipitation seasonality, cation exchange capacity of the soil, and soil pH. The four SNPs correlated with this PC might be involved in soil-related adaptation. The minor allele frequencies (MAFs) of the 10 significant SNPs were correlated with the environmental PCs (Figures 3 and S4). The absolute correlation coefficients varied between 0.34 and 0.65 (mean ϱ: 0.52). This observed pattern is a clear sign of local environmental adaptation of the investigated European beech populations and might be considered in future tree breeding programs (see below).

The association analysis between SNPs and tree traits revealed 12 significant SNPs that were all associated with SLA. SLA [or its inverse value leaf mass per area (LMA); Poorter et al., 2009] was shown to vary along precipitation gradients (Poorter et al., 2009; Salehi et al., 2020; Wright et al., 2004). Usually, SLA is lower in dry environments, but an opposite trend was found for European beech. Along precipitation gradients in Germany and Switzerland, SLA increased with decreasing mean annual precipitation (Meier & Leuschner, 2008; Salehi et al., 2020). However, for the beech populations in the present study, it was shown that SLA decreased with a reduction in soil water storage capacity, suggesting that leaf development in beech is primarily constrained by the soil water pool and less by the amount of rainfall (Weithmann et al., 2022c).

We used pRDA-based variance partitioning to infer the relative contribution of traits, neutral population structure, and geography to the genetic variation. The full model, including all variables, explained

8.36% of the total variance. Traits alone explained 5.97% of the total variance and 71.37% of the explainable variance. Both neutral population structure and geography had smaller effects and explained 0.97% and 1.04% of the total variance (11.6% and 12.47% of the full model). Confounding factors (i.e., variance that cannot be uniquely attributed to any of the three variables) explained only 0.38% of the total variance (4.56% of the full model). Thus, most of the explainable genetic variation can be attributed to the analyzed traits. This means that these morphological and tree hydraulic traits could successfully be altered in breeding programs to enhance the adaptation of European beech to changing environmental conditions. Thereby, SNP sets, like the one presented in this study, may be used for marker-based selection of seedlings with desired traits. Genomic selection, in which genetic markers distributed throughout the genome of a species are used to train models and select suitable genotypes in breeding programs, is being applied more and more in forest tree species and might soon become the most efficient method in tree breeding (Grattapaglia, 2022).

It is expected that associations in complex traits are spread across most of the genome (Boyle et al., 2017). This was also the case in the present study. The significant SNPs were distributed over 8 of the 12 European beech chromosomes (Table 3) (Mishra et al., 2022). Several genes, in which significantly associated SNPs were located, are related to stress response. For instance, serine hydroxymethyltransferases (intronic SNP associated with PC2) have been shown to be involved in the response to environmental stresses in different plant species (Fang et al., 2020; Liu et al., 2022; Nogués et al., 2022). Eukaryotic translational initiation factor 4A (intronic SNP associated with SLA) belongs to the family of helicases. They can stimulate stress-induced pathways and have been shown to increase stress tolerance in groundnut (Santosh Rama Bhadra Rao et al., 2017; Tuteja et al., 2014). Finally, ubiquitin carboxyl-terminal hydrolases (intronic SNP associated with SLA) were found to be needed for the period maintenance of the circadian clock during high temperatures in *Arabidopsis* (Hayama et al., 2019). In total, 18 of the 22 significant SNPs were non-coding SNPs (9 intergenic and 9 intronic SNPs). Especially, the overrepresentation of intergenic SNPs (8.1% intergenic SNPs in the total SNP set and 41% intergenic SNPs among the significant SNPs) is surprising since they should not be directly involved in trait expression. One possible explanation would be that they are in linkage disequilibrium with the causative SNPs. One of the significant intergenic SNPs was located 248 bp away from the next gene (*Casein kinase II subunit alpha-2*). Potentially, the SNP is located within the promoter region of the gene and involved in the regulation of gene expression, but this needs further validation (Klees et al., 2022; Wittkopp & Kalay, 2012).

Our approach to only consider SNPs as significant when they were consistently detected with different association methods might be quite conservative. This approach is often used to reduce the number of false positive associations but, at the same time, true associations (false negatives) might be lost (Storfer et al., 2018). When considering all SNPs detected with at least one of the applied association methods in the environmental association analysis, a total of 300 significant SNPs were detected. Likely, not all of these SNPs are false positives, and hence, we provide them for comparisons with future studies in the supplementary material of this publication (data file S3).

CONCLUSIONS

The new SNP set for European beech revealed by our study can be used for future research in this species. We identified SNPs and candidate genes that are potentially involved in the control of specific leaf area (SLA) and general environmental adaptation in beech. Future studies may be conducted to confirm these associations in independent populations. Clear patterns of adaptation to local environmental conditions were detected in the beech populations. Further, the analyzed morphological and hydraulic traits explained most of the explainable genetic variation. Thus, they could successfully be altered in tree breeding programs, which may help to increase the adaptation of European beech to changing environmental conditions. Similar studies with European beech and other tree species are required to deepen our knowledge of the genetic basis of tree physiological traits. In general, there is a strong need for a collaboration between genomics and plant physiology to address the most pressing challenges in the field of plant sciences (Interdisciplinary Plant Science Consortium, 2023).

AUTHOR CONTRIBUTIONS

Markus Müller: conceptualization; data curation; formal analysis; investigation; methodology; validation; visualization; writing – original draft preparation. **Christoph Leuschner:** conceptualization; funding acquisition; resources; project administration; supervision; writing – review & editing. **Greta Weithmann:** data curation; formal analysis; investigation; resources; writing – review & editing. **Robert Weigel:** data curation; formal analysis; investigation; resources; visualization; writing – review & editing. **Bat-Enerel Banzragch:** data curation; formal analysis; investigation; resources; writing – review & editing. **Wilfried Steiner:** conceptualization; funding acquisition; writing – review & editing. **Oliver Gailing:** conceptualization; funding acquisition; project administration; resources; supervision; writing – review & editing.

ACKNOWLEDGEMENTS

We thank Katharina Birgit Budde and Andrii Tarieiev for assistance with data analysis and Alexandra Dolynska for help with laboratory work. The project was funded by the German Federal Ministry of Food and Agriculture (BMEL) and the German Federal Ministry for the Environment, Nature Conservation, Nuclear Safety, and Consumer Protection (BMUV) represented by the Fachagentur Nachwachsende Rohstoffe (FNR) within the framework of the Waldklimafonds (projects BEECHLIMITS (reference number: 22WC415001) and GenVar-Buche (reference number: 2218WK34A4). We thank the managers of the three Exploratories, Kirsten Reichel-Jung, Florian Staub, Juliane Vogt, Anna K. Franke, Miriam Teuscher, Franca Marian and all former managers for their work in maintaining the plot and project

infrastructure; Christiane Fischer and Jule Mangels for giving support through the central office, Andreas Ostrowski for managing the central data base, and Markus Fischer, Eduard Linsenmair, Dominik Hessenmöller, Daniel Prati, Ingo Schöning, François Buscot, Ernst-Detlef Schulze, Wolfgang W. Weisser and the late Elisabeth Kalko for their role in setting up the Biodiversity Exploratories project. We thank the administration of the Hainich national park, the UNESCO Biosphere Reserve Swabian Alb and the UNESCO Biosphere Reserve Schorfheide-Chorin as well as all land owners for the excellent collaboration. The work has been partly funded by the DFG Priority Program 1374 "Biodiversity-Exploratories" (GENEDIV, project number 433288081). Field work permits were issued by the responsible state environmental offices of Baden-Württemberg, Thüringen, and Brandenburg. Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

The project was funded by the German Federal Ministry of Food and Agriculture (BMEL) and the German Federal Ministry for the Environment, Nature Conservation, Nuclear Safety, and Consumer Protection (BMUV) represented by the Fachagentur Nachwachsende Rohstoffe (FNR) within the framework of the Waldklimafonds (projects BEECHLIMITS and GenVarBuche) [grant numbers: 22WC415001; 2218WK34A4]; the work has been partly funded by the DFG Priority Program 1374 "Biodiversity-Exploratories" (project GENEDIV) [grant number: 433288081].

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in NCBI SRA at https://www.ncbi.nlm.nih.gov/sra, reference number PRJNA1043878.

ORCID

Markus Müller https://orcid.org/0000-0001-9990-0719
Christoph Leuschner https://orcid.org/0000-0002-5689-7932
Greta Weithmann https://orcid.org/0000-0001-5876-698X
Robert Weigel https://orcid.org/0000-0001-9685-6783
Bat-Enerel Banzragch https://orcid.org/0000-0003-3678-5308
Oliver Gailing https://orcid.org/0000-0002-4572-2408

REFERENCES

Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T., & Curtis-McLane, S. (2008). Adaptation, migration or extirpation: Climate change outcomes for tree populations. *Evolutionary Applications*, 1(1), 95–111. https://doi.org/10.1111/j.1752-4571.2007.00013.x

Arvidsson, S., Fartmann, B., Winkler, S., & Zimmermann, W. (2016). *Efficient high-throughput SNP discovery and genotyping using normalised Genotyping-by-Sequencing (nGBS)*. LGC Technical Note, AN-161104.01.

Beaudette, D., Skovlin, J., Roecker, S., & Brown, A. (2022). soilDB: soil database interface. *R package version* 2 6. 13.

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

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: From polygenic to omnigenic. *Cell*, 169(7), 1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Capblancq, T., & Forester, B. R. (2021). Redundancy analysis: A swiss army knife for landscape genomics. *Methods in Ecology and Evolution*, 12(12), 2298–2309. https://doi.org/10.1111/2041-210X.13722

Capblancq, T., Luu, K., Blum, M. G. B., & Bazin, E. (2018). Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources*, 18(6), 1223–1233. https://doi.org/10.1111/1755-0998.12906

Capblancq, T., Morin, X., Gueguen, M., Renaud, J., Lobreaux, S., & Bazin, E. (2020). Climate-associated genetic variation in *Fagus sylvatica* and potential responses to climate change in the French Alps. *Journal of Evolutionary Biology*, 33(6), 783–796. https://doi.org/10.1111/jeb.13610

Chhatre, V. E., & Emerson, K. J. (2017). StrAuto: Automation and parallelization of STRUCTURE analysis. *BMC Bioinformatics*, 18(1), 192. https://doi.org/10.1186/s12859-017-1593-0

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012a). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6(2), 80–92. https://doi.org/10.4161/fly.19695

Cingolani, P., Patel, V., Coon, M., Nguyen, T., Land, S., Ruden, D., & Lu, X. (2012b). Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Frontiers in Genetics*, 3, 35. https://doi.org/10.3389/fgene.2012.00035

Csilléry, K., Lalagüe, H., Vendramin, G. G., González-Martínez, S. C., Fady, B., & Oddou-Muratorio, S. (2014). Detecting short spatial scale local adaptation and epistatic selection in climate-related candidate genes in European beech (*Fagus sylvatica*) populations. *Molecular Ecology*, 23(19), 4696–4708. https://doi.org/10.1111/mec.12902

Cuervo-Alarcon, L., Arend, M., Müller, M., Sperisen, C., Finkeldey, R., & Krutovsky, K. V. (2018). Genetic variation and signatures of natural selection in populations of European beech (*Fagus sylvatica* L.) along precipitation gradients. *Tree Genetics & Genomes*, 14(6), 84. https://doi.org/10.1007/s11295-018-1297-2

Cuervo-Alarcon, L., Arend, M., Müller, M., Sperisen, C., Finkeldey, R., & Krutovsky, K. V. (2021). A candidate gene association analysis identifies SNPs potentially involved in drought tolerance in European beech (*Fagus sylvatica* L.). *Scientific Reports*, 11(1), 2386. https://doi.org/10.1038/s41598-021-81594-w

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Ellenberg, H. (1963). *Vegetation Mitteleuropas mit den Alpen*. Eugen Ulmer.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Fang, C., Zhang, P., Li, L., Yang, L., Mu, D., Yan, X., Li, Z., & Lin, W. (2020). Serine hydroxymethyltransferase localised in the endoplasmic reticulum plays a role in scavenging H_2O_2 to enhance rice chilling tolerance. *BMC Plant Biology*, 20(1), 236. https://doi.org/10.1186/s12870-020-02446-9

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302–4315. https://doi.org/10.1002/joc.5086

Frichot, E., & François, O. (2015). LEA: an R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6(8), 925–929. https://doi.org/10.1111/2041-210x.12382

Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent

factor mixed models. *Molecular Biology and Evolution*, 30(7), 1687–1699. https://doi.org/10.1093/molbev/mst063

García-Plazaola, J. I., & Becerril, J. M. (2000). Effects of drought on photoprotective mechanisms in European beech (*Fagus sylvatica* L.) seedlings from different provenances. *Trees*, 14(8), 485–490. https://doi.org/10.1007/s004680000068

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv*, *arXiv: 1207.3907*.

Grattapaglia, D. (2022). Twelve years into genomic selection in forest trees: Climbing the slope of enlightenment of marker assisted tree breeding. *Forests*, 13(10), 1554. https://doi.org/10.3390/f13101554

Gezan, S., de Oliveira, A.A., Galli, G., & Murray, D. (2022). ASRgenomics: An R package with complementary genomic functions. Version 1.1.0 VSN International, Hemel Hempstead, United Kingdom.

Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). dartR: An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources*, 18(3), 691–699. https://doi.org/10.1111/1755-0998.12745

Günther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics*, 195(1), 205–220. https://doi.org/10.1534/genetics.113.152462

Hayama, R., Yang, P., Valverde, F., Mizoguchi, T., Furutani-Hayama, I., Vierstra, R. D., & Coupland, G. (2019). Ubiquitin carboxyl-terminal hydrolases are required for period maintenance of the circadian clock at high temperature in *Arabidopsis*. *Scientific Reports*, 9(1), 17030. https://doi.org/10.1038/s41598-019-53229-8

Hijmans, R. J. (2022). Raster: Geographic data analysis and modeling. *R package version* 3. 5–15.

Interdisciplinary Plant Science Consortium. (2023). Inclusive collaboration across plant physiology and genomics: Now is the time! *Plant Direct*, 7(5), e493. https://doi.org/10.1002/pld3.493

Kamvar ZN, Tabima JF, & Grünwald NJ (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281. https://doi.org/10.7717/peerj.281

Kamvar ZN, Brooks JC, & Grünwald NJ (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, 6, 208. https://doi.org10.3389/fgene.2015.00208

Karabayev, D., Molkenov, A., Yerulanuly, K., Kabimoldayev, I., Daniyarov, A., Sharip, A., Seisenova, A., Zhumadilov, Z., & Kairov, U. (2021). re-Searcher: GUI-based bioinformatics tool for simplified genomics data mining of VCF files. *PeerJ*, 9, e11333. https://doi.org/10.7717/peerj.11333

Klees, S., Heinrich, F., Schmitt, A. O., & Gültas, M. (2022). agReg-SNPdb-Plants: A database of regulatory SNPs for agricultural plant species. *Biology*, 11(5), 684. https://doi.org/10.3390/biology11050684

Knaus, B. J., & Grünwald, N. J. (2017). vcfr: A package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, 17(1), 44–53. https://doi.org/10.1111/1755-0998.12549

Knutzen, F., Dulamsuren, C., Meier, I. C., & Leuschner, C. (2017). Recent climate warming-related growth decline impairs European beech in the center of its distribution range. *Ecosystems*, 20(8), 1494–1511. https://doi.org/10.1007/s10021-017-0128-x

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across *K*. *Molecular Ecology Resources*, 15(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

Krajmerová, D., Hrivnák, M., Ditmarová, L., Jamnická, G., Kmeť, J., Kurjak, D., & Gömöry, D. (2017). Nucleotide polymorphisms associated with climate, phenology and physiological traits in European beech (*Fagus sylvatica* L.). *New Forests*, 48(3), 463–477. https://doi.org/10.1007/s11056-017-9573-9

Lalagüe, H., Csilléry, K., Oddou-Muratorio, S., Safrana, J., de Quattro, C., Fady, B., González-Martínez, S. C., & Vendramin, G. G. (2014). Nucleotide diversity and linkage disequilibrium at 58 stress response and phenology candidate genes in a European beech (*Fagus sylvatica* L.) population from southeastern France. *Tree Genetics & Genomes*, 10(1), 15. https://doi.org/10.1007/s11295-013-0658-0

Le Corre, V., & Kremer, A. (2012). The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology Notes*, 21(7), 1548–1566. https://doi.org/10.1111/j.1365-294X.2012.05479.x

Leuschner, C. (2020). Drought response of European beech (*Fagus sylvatica* L.)—A review. *Perspectives in Plant Ecology, Evolution and Systematics*, 47, 125576. https://doi.org/10.1016/j.ppees.2020.125576

Lind, B. M., Menon, M., Bolte, C. E., Faske, T. M., & Eckert, A. J. (2018). The genomics of local adaptation in trees: Are we out of the woods yet? *Tree Genetics & Genomes*, 14(2), 29. https://doi.org/10.1007/s11295-017-1224-y

Liu, H., Li, N., Zhao, Y., Kang, G.-Z., Zhao, Y.-H., & Xu, H.-W. (2022). Serine hydroxymethyltransferase (*SHMT*) gene family in wheat (*Triticum aestivum* L.): Identification, evolution, and expression analysis. *Agronomy*, 12(6), 1346. https://doi.org/10.3390/agronomy12061346

Martinez del Castillo, E., Zang, C. S., Buras, A., Hacket-Pain, A., Esper, J., Serrano-Notivoli, R., Hartl, C., Weigel, R., Klesse, S., Resco de Dios, V., Scharnweber, T., Dorado-Liñán, I., van der Maaten-Theunissen, M., van der Maaten, E., Jump, A., Mikac, S., Banzragch, B.-E., Beck, W., Cavin, L., … de Luis, M. (2022). Climate-change-driven growth decline of European beech forests. *Communications Biology*, 5(1), 163. https://doi.org/10.1038/s42003-022-03107-3

Meger, J., Ulaszewski, B., & Burczyk, J. (2021). Genomic signatures of natural selection at phenology-related genes in a widely distributed tree species *Fagus sylvatica* L. *BMC Genomics*, 22(1), 583. https://doi.org/10.1186/s12864-021-07907-5

Meier, I. C., & Leuschner, C. (2008). Leaf size and leaf area index in *Fagus sylvatica* forests: Competing effects of precipitation, temperature, and nitrogen availability. *Ecosystems*, 11(5), 655–669. https://doi.org/10.1007/s10021-008-9135-2

Mishra, B., Gupta, D. K., Pfenninger, M., Hickler, T., Langer, E., Nam, B., Paule, J., Sharma, R., Ulaszewski, B., Warmbier, J., Burczyk, J., & Thines, M. (2018). A reference genome of the European beech (*Fagus sylvatica* L.). *Gigascience*, 7(6), giy063. https://doi.org/10.1093/gigascience/giy063

Mishra, B., Ulaszewski, B., Meger, J., Aury, J.-M., Bodénès, C., Lesur-Kupin, I., Pfenninger, M., Da Silva, C., Gupta, D. K., Guichoux, E., Heer, K., Lalanne, C., Labadie, K., Opgenoorth, L., Ploch, S., Le Provost, G., Salse, J., Scotti, I., Wötzel, S., … Thines, M. (2022). A chromosome-level genome assembly of the European beech (*Fagus sylvatica*) reveals anomalies for organelle DNA integration, repeat content and distribution of SNPs. *Frontiers in Genetics*, 12, 691058. https://doi.org/10.3389/fgene.2021.691058

Müller, M., Cuervo-Alarcon, L., Gailing, O., K. C, R., Chhetri, M., Seifert, S., Arend, M., Krutovsky, K., & Finkeldey, R. (2018). Genetic variation of European beech populations and their progeny from northeast Germany to southwest Switzerland. *Forests*, 9(8), 469. https://doi.org/10.3390/f9080469

Müller, M., Seifert, S., & Finkeldey, R. (2015). A candidate gene-based association study reveals SNPs significantly associated with bud burst in European beech (*Fagus sylvatica* L.). *Tree Genetics & Genomes*, 11(6), 116. https://doi.org/10.1007/s11295-015-0943-1

Müller, M., Seifert, S., & Finkeldey, R. (2017). Comparison and confirmation of SNP-bud burst associations in European beech populations in Germany. *Tree Genetics & Genomes*, 13, 59. https://doi.org/10.1007/s11295-017-1145-9

Neycken, A., Scheggia, M., Bigler, C., & Lévesque, M. (2022). Long-term growth decline precedes sudden crown dieback of European beech. *Agricultural and Forest Meteorology*, 324, 109103. https://doi.org/10.1016/j.agrformet.2022.109103

Nogués, I., Sekula, B., Angelaccio, S., Grzechowiak, M., Tramonti, A., Contestabile, R., & Ruszkowski, M. (2022). *Arabidopsis thaliana* serine

hydroxymethyltransferases: Functions, structures, and perspectives. *Plant Physiology and Biochemistry*, 187, 37–49. https://doi.org/10.1016/j.plaphy.2022.07.025

NW-FVA. (2019). Komplexe Schäden an Rotbuche (*Fagus sylvatica*) und Auswirkungen des trockenen und heißen Sommers 2018 auf ältere Bestände. *Waldschutzinfo Nr. 06/2019*.

Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2020). Vegan: Community ecology package. *R package version* 2. 5–7.

Peuke, A. D., Schraml, C., Hartung, W., & Rennenberg, H. (2002). Identification of drought-sensitive beech ecotypes by physiological parameters. *New Phytol*, 154(2), 373–387. https://doi.org/10.1046/j.1469-8137.2002.00400.x

Pfenninger, M., Langan, L., Feldmeyer, B., Fussi, B., Hoffmann, J., Granado, R., Hetzer, J., Šeho, M., Mellert, K.-H., & Hickler, T. (2023). *Phenotypic drought stress prediction of European beech* (Fagus sylvatica) *by genomic prediction and remote sensing* [Preprint]. *Evolutionary Biology*. https://doi.org/10.1101/2023.03.29.534688

Pfenninger, M., Reuss, F., Kiebler, A., Schönnenbeck, P., Caliendo, C., Gerber, S., Cocchiararo, B., Reuter, S., Blüthgen, N., Mody, K., Mishra, B., Bálint, M., Thines, M., & Feldmeyer, B. (2021). Genomic basis for drought resistance in European beech forests threatened by climate change. *Elife*, 10, e65532. https://doi.org/10.7554/eLife.65532

Pluess, A. R., & Weber, P. (2012). Drought-adaptation potential in *Fagus sylvatica*: Linking moisture availability with genetic diversity and dendrochronology. *PLoS One*, 7(3), e33636. https://doi.org/10.1371/journal.pone.0033636

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *Soil*, 7(1), 217–240. https://doi.org/10.5194/soil-7-217-2021

Poorter, H., Niinemets, Ü., Poorter, L., Wright, I. J., & Villar, R. (2009). Causes and consequences of variation in leaf mass per area (LMA): A meta-analysis. *New Phytologist*, 182(3), 565–588. https://doi.org/10.1111/j.1469-8137.2009.02830.x

Postolache, D., Oddou-Muratorio, S., Vajana, E., Bagnoli, F., Guichoux, E., Hampe, A., Le Provost, G., Lesur, I., Popescu, F., Scotti, I., Piotti, A., & Vendramin, G. G. (2021). Genetic signatures of divergent selection in European beech (*Fagus sylvatica* L.) are associated with the variation in temperature and precipitation across its distribution range. *Molecular Ecology*, 30(20), 5029–5047. https://doi.org/10.1111/mec.16115

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959.

R Core Team. (2021). *R: A language and environment for statistical computing*.

Revelle, W. (2021). psych: Procedures for personality and psychological research. https://CRAN.R-project.org/package=psych

Rose, L., Leuschner, C., Köckemann, B., & Buschmann, H. (2009). Are marginal beech (*Fagus sylvatica* L.) provenances a source for drought tolerant ecotypes? *European Journal of Forest Research*, 128(4), 335–343. https://doi.org/10.1007/s10342-009-0268-4

Salehi, M., Walthert, L., Zimmermann, S., Waldner, P., Schmitt, M., Schleppi, P., Liechti, K., Ahmadi, M., Zahedi Amiri, G., Brunner, I., & Thimonier, A. (2020). Leaf morphological traits and leaf nutrient concentrations of European beech across a water availability gradient in Switzerland. *Frontiers in Forests and Global Change*, 3, 19. https://doi.org/10.3389/ffgc.2020.00019

Santosh Rama Bhadra Rao, T., Vijaya Naresh, J., Sudhakar Reddy, P., Reddy, M. K., & Mallikarjuna, G. (2017). Expression of *Pennisetum glaucum* eukaryotic translational initiation factor 4A (*PgeIF4A*) confers improved drought, salinity, and oxidative stress tolerance in groundnut. *Frontiers in Plant Science*, 8, 453. https://doi.org/10.3389/fpls.2017.00453

Schönnenbeck, P., Schell, T., Gerber, S., & Pfenninger, M. (2021). tbg—A new file format for genomic data. *bioRxiv*. https://doi.org/10.1101/2021.03.15.435393

Storfer, A., Patton, A., & Fraik, A. K. (2018). Navigating the interface between landscape genetics and landscape genomics. *Frontiers in Genetics*, 9, 68. https://doi.org/10.3389/fgene.2018.00068

Tuteja, N., Banu, Mst. S. A., Huda, K. Md. K., Gill, S. S., Jain, P., Pham, X. H., & Tuteja, R. (2014). Pea p68, a DEAD-box helicase, provides salinity stress tolerance in transgenic tobacco by reducing oxidative stress and improving photosynthesis machinery. *PLoS ONE*, 9(5), e98287. https://doi.org/10.1371/journal.pone.0098287

Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *Journal of Climate*, 23(7), 1696–1718. https://doi.org/10.1175/2009JCLI2909.1

Wang, J., & Zhang, Z. (2021). GAPIT version 3: Boosting power and accuracy for genomic association and prediction. *Genomics, Proteomics & Bioinformatics*, 19(4), 629–640. https://doi.org/10.1016/j.gpb.2021.08.005

Weigel, R., Bat-Enerel, B., Dulamsuren, C., Muffler, L., Weithmann, G., & Leuschner, C. (2023). Summer drought exposure, stand structure, and soil properties jointly control the growth of European beech along a steep precipitation gradient in northern Germany. *Global Change Biology*, 29, 763–779. https://doi.org/10.1111/gcb.16506

Weithmann, G., Link, R. M., Banzragch, B. E., Würzberg, L., Leuschner, C., & Schuldt, B. (2022a). Soil water availability and branch age explain variability in xylem safety of European beech in Central Europe. *Oecologia*, 198(3), 629–644. https://doi.org/10.1007/s00442-022-05124-9

Weithmann, G., Paligi, S. S., Schuldt, B., & Leuschner, C. (2022b). Branch xylem vascular adjustments in European beech in response to decreasing water availability across a precipitation gradient. *Tree Physiology*, 42(11), 2224–2238. https://doi.org/10.1093/treephys/tpac080

Weithmann, G., Schuldt, B., Link, R. M., Heil, D., Hoeber, S., John, H., Müller-Haubold, H., Schüller, L. M., Schumann, K., & Leuschner, C. (2022c). Leaf trait modification in European beech trees in response to climatic and edaphic drought. *Plant Biology*, 24(7), 1272–1286. https://doi.org/10.1111/plb.13366

Wittkopp, P. J., & Kalay, G. (2012). *Cis*-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, 13(1), 59–69. https://doi.org/10.1038/nrg3095

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J. H. C., Diemer, M., Flexas, J., Garnier, E., Groom, P. K., Gulias, J., Hikosaka, K., Lamont, B. B., Lee, T., Lee, W., Lusk, C., … Villar, R. (2004). The worldwide leaf economics spectrum. *Nature*, 428(6985), 821–827. https://doi.org/10.1038/nature02403

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Müller, M., Leuschner, C., Weithmann, G., Weigel, R., Banzragch, B.-E., Steiner, W. et al. (2024) A genome-wide genetic association study reveals SNPs significantly associated with environmental variables and specific leaf area in European beech. *Physiologia Plantarum*, 176(3), e14334. Available from: https://doi.org/10.1111/ppl.14334