

# 1 Genomic basis of drought resistance in 2 *Fagus sylvatica*

3  
4 Markus Pfenninger<sup>1,2,3</sup>, Friederike Reuss<sup>1</sup>, Angelika Kiebler<sup>1</sup>, Philipp Schönnenbeck<sup>4</sup>,  
5 Cosima Caliendo<sup>1,4</sup>, Susanne Gerber<sup>4</sup>, Berardino Cocchiararo<sup>3,5</sup>, Sabrina Reuter<sup>6</sup>, Nico  
6 Blüthgen<sup>6</sup>, Karsten Mody<sup>6,11</sup>, Bagdevi Mishra<sup>7</sup>, Miklós Bálint<sup>3,8,9</sup>, Marco Thines<sup>3,7,10</sup>,  
7 Barbara Feldmeyer<sup>1</sup>

8  
9 <sup>1</sup> Molecular Ecology, Senckenberg Biodiversity and Climate Research Centre, Frankfurt am Main,  
10 Germany

11 <sup>2</sup> Institute for Organismic and Molecular Evolution, Johannes Gutenberg University, Mainz, Germany

12 <sup>3</sup> LOEWE Centre for Translational Biodiversity Genomics, Frankfurt am Main, Germany

13 <sup>4</sup> Institute of Human Genetics, University Medical Center, Johannes Gutenberg University, Mainz,  
14 Germany

15 <sup>5</sup> Conservation Genetics Section, Senckenberg Research Institute and Natural History Museum  
16 Frankfurt, Gelnhausen, Germany

17 <sup>6</sup> Ecological Networks lab, Department of Biology, Technische Universität Darmstadt, Darmstadt,  
18 Germany

19 <sup>7</sup> Biological Archives, Senckenberg Biodiversity and Climate Research Centre, Frankfurt am Main,  
20 Germany

21 <sup>8</sup> Functional Environmental Genomics, Senckenberg Biodiversity and Climate Research Centre,  
22 Frankfurt am Main, Germany

23 <sup>9</sup> Agricultural Sciences, Nutritional Sciences, and Environmental Management, Universität Giessen,  
24 Giessen, Germany

25 <sup>10</sup> Institute for Ecology, Evolution and Diversity, Johann Wolfgang Goethe-University, Frankfurt am  
26 Main, Germany

27 <sup>11</sup> Department of Applied Ecology, Hochschule Geisenheim University, Geisenheim, Germany

## 28 **Abstract**

29 In the course of global climate change, central Europe is experiencing more frequent and prolonged  
30 periods of drought. The drought years 2018 and 2019 affected European beeches (*Fagus sylvatica* L.)  
31 differently: even in the same stand, drought damaged trees neighboured healthy trees, suggesting  
32 that the genotype rather than the environment was responsible for this conspicuous pattern. We  
33 used this natural experiment to study the genomic basis of drought resistance with Pool-GWAS.  
34 Contrasting the extreme phenotypes identified 106 significantly associated SNPs throughout the  
35 genome. Most annotated genes with associated SNPs (>70%) were previously implicated in the  
36 drought reaction of plants. Non-synonymous substitutions led either to a functional amino acid  
37 exchange or premature termination. A SNP-assay with 70 loci allowed predicting drought phenotype

38 in 98.6% of a validation sample of 92 trees. Drought resistance in European beech is a moderately  
39 polygenic trait that should respond well to natural selection, selective management, and breeding.

## 40 **Keywords**

41 Genome-wide association study, genomic prediction, forest tree, Fagales, conservation genomics,  
42 functional environmental genomics

## 43 **Impact Statement**

44 European beech harbours substantial genetic variation at genomic loci associated with drought  
45 resistance and the loci identified in this study can help to accelerate and monitor adaptation to  
46 climate change.

## 47 **Introduction**

48 Climate change comes in many different facets, amongst which are prolonged drought periods  
49 (Christensen *et al.* 2007). The Central European droughts in the years 2018 and 2019 caused severe  
50 water stress in many forest tree species, leading to the die-off of many trees (Schuldt *et al.* 2020a).  
51 Among the suffering tree species was European beech, *Fagus sylvatica* L. As one of the most  
52 common deciduous tree species in Central Europe, *F. sylvatica* is of great ecological importance:  
53 beech forests are a habitat for more than 6,000 different animal and plant species (Brunet *et al.*  
54 2010; Dorow *et al.* 2010). The forestry use of beech in 2017 generated a turnover of more than 1  
55 billion € in Germany alone (Thünen\_Institute 2020), without taking the economic and societal value  
56 of the ecosystem services of woods into account (Elsasser *et al.* 2016). However, the drought years  
57 2018 and 2019 severely impacted the beech trees in Germany (Paar & Dammann 2019). Official  
58 reports on drought damage in beech recorded 62% of trees with rolled leaves and 20-30% of small  
59 leaves, mainly in the crown, resulting in 7% of badly damaged or dead trees. As shown before  
60 (Bressemer 2008), most trees affected by drought stress were medium to old aged.

61 Under favourable conditions, beech is a competitive and shade tolerant tree species, dominating  
62 mixed stands (Pretzsch *et al.* 2013). High genetic diversity within populations supports adaptation to  
63 local conditions (Kreyling *et al.* 2012). Significant differences between local populations in tolerance  
64 to various stress factors such as early frost (Czajkowski & Bolte 2006), drought (Cocozza *et al.* 2016;  
65 Harter *et al.* 2015) or air pollution (Müller-Starck 1985) are known. The distribution of *F. sylvatica* is  
66 mainly limited by water-availability, as the tree does not tolerate particularly wet or dry conditions  
67 (Sutmöller *et al.* 2008). Therefore, it is quite conceivable that the species could suffer even more  
68 under the predicted future climatic conditions than today (Sutmöller *et al.* 2008).

69 Despite the widespread, severe drought damage, a pattern observed in all beech forests was very  
70 noticeable (personal observations). Using crown deterioration as significant indicator for drought  
71 damage (Choat *et al.* 2018), not all trees in a beech stand were equally damaged or healthy. The  
72 damage occurred rather in a mosaic-like pattern instead. Even though the extent of drought damage  
73 varied among sites, apparently completely healthy trees immediately neighboured severely damaged  
74 ones and *vice versa*. This observation gave rise to the hypothesis that not the local environmental  
75 conditions might be decisive for the observed drought damage, but rather the genetic make-up of  
76 the individual trees.

77 We decided to draw on this natural “experimental set-up” to infer the genomic basis underlying the  
78 drought susceptibility in *F. sylvatica*. We identified more than 200 neighbouring pairs of trees with  
79 extreme phenotypes and used a Pool-GWAS approach (Bastide *et al.* 2013) to infer associated SNP  
80 loci by contrasting allele frequencies with replicated pools of drought susceptible and resistant  
81 individuals. In addition, we individually re-sequenced a subset of 51 pairs of susceptible and resistant  
82 trees. If the observed pattern indeed has a genetic basis, identifying the associated loci would enable

83 the genomic prediction of drought resistance (Stocks *et al.* 2019). Constructing a SNP assay from the  
84 most highly phenotype associated SNPs, we validated 70 identified loci by predicting the drought  
85 phenotype of an additional set of beech trees from their genotype at these loci using Linear  
86 Discriminant Analysis and a new Machine Learning approach (Horenko 2020). These accurate  
87 genomic prediction tools, e.g., the choice of drought resistant seed producing trees and selective  
88 logging could help accelerate and monitor natural selection and thus harness beech forests against  
89 climate change (Waldvogel *et al.* 2020).

## 90 Results

### 91 *Sampling, climate development and phenotyping*

92 Damaged and healthy beech tree pairs were sampled from woods in the lowland Rhein-Main plain,  
93 the adjacent low mountain ranges of Odenwald and Taunus, and mountain ranges from Central and  
94 Northern Hessen (Fig. 1A). When summarising the climatic conditions from 1950 to 2019 for the  
95 sampling sites in a principal component analysis (PCA), the sites were divided into two groups by axis  
96 1, a temperature gradient. The Taunus mountain sites grouped with those from the northern part of  
97 Hessen, while the Rhein-Main plain clustered with the Odenwald sites (Fig. 1B). This grouping was  
98 also used to construct the GWAS pools (see below). Comparing the climate from the 1950s, when  
99 most of the trees sampled were already in place, with the decade from 2010-2019, showed that all  
100 local conditions changed substantially and similarly in the direction and extent of warmer and drier  
101 conditions (Fig. 1B). The steepest temperature increase occurred in the 1980s, while precipitation  
102 patterns mainly changed in the last decade (Suppl. Fig. 1). A wide range of parameters, potentially  
103 relevant as selection pressures changed drastically during this period: the mean January daily  
104 minimum temperature at the sampling sites increased by 1.49°C from -2.64°C (s.d. 1.68°C) in the  
105 1950s to -1.15°C (s.d. 2.50°C) during the last decade. The mean August daily maximum temperatures  
106 increased even more by 2.37°C from 22.06°C (s.d. 1.95°C) to 24.43°C (s.d. 2.35°C). Simultaneously,  
107 mean annual precipitation decreased by 40.5 mm or 5.5% from 741.2 mm (s.d. 85.8 mm) to 700.7  
108 mm (s.d. 70.9 mm). Most of the precipitation loss (84%) occurred during the main growth period  
109 between April and September, with a decrease of 33.9 mm from 410.4 mm (s.d. 36.1 mm) to 376.5  
110 mm (s.d. 25.6 mm).

111 Mean monthly evaporation potential, available from 1991 onwards, showed that, compared to the  
112 beginning of the 1990s, the main growth period of beech from April to September became  
113 increasingly drier, with up to 30 mm more evaporation per month. The drought dynamics suggested  
114 that the years 2018 and 2019 were not outliers, but rather part of a long-term, accelerating trend  
115 (Fig. 1C), following the overall global pattern (Büntgen *et al.* 2021; Trenberth *et al.* 2014).

116 There was a strong negative correlation ( $r = 0.695$ ) between the drought strength during the main  
117 growth period (Apr- Sept) and a proxy for (green) leaf cover (leaf area index, LAI) for the sampled  
118 plots in the years 2015-2019 (Suppl. Fig. 2). This observation suggested that leaf loss and dried leaves  
119 are good indicators for drought stress.

120 The mean distance between paired trees was 5.1 m (s.d. 3.4 m, Suppl. Fig. 3). Phenotypic  
121 measurements generally confirmed the study design and selection of trees: healthy and damaged  
122 trees within each tree pair did not differ significantly in *trunk circumference*, *tree height*, *canopy*  
123 *closure* and *competition index* (Fig. 2 A-D, Suppl. Table 2). Hence, these parameters were not  
124 considered in further analyses. As expected, and confirming the assignment of damage status, the  
125 quantity of *dried leaves* and *leaf loss* differed substantially between damaged and healthy ones (Fig.  
126 2 E-F, Suppl. Table 2). A sample of photographs contrasting damaged and healthy paired trees can be  
127 found in the Suppl. Fig. 4.

## 128 *Linkage disequilibrium, population structure and genome-wide association study*

129 For a subsample of 300 out of the 402 sampled beech trees we generated four DNA pools from two  
130 climatically distinct regions (North and South Hessen, Fig. 1B), contrasting trees that were either  
131 healthy or highly drought damaged respectively (Tab. S1). The “South” pools consisted of 100  
132 individuals each, whereas the “North” pools contained 50 individuals each. We created ~50GB 150  
133 bp-paired end reads with insert size 250-300 bp on an Illumina HiSeq 4000 system per pool. More  
134 than 96% of the reads mapped against the repeat-masked chromosome level beech reference  
135 genome (accession no. PRJNA450822). After filtering the alignment for quality and a coverage  
136 between 15x and 70x, and removing indels, allele frequencies for 9.6 million SNPs were scored. All  
137 100 individuals from the North population were additionally individually re-sequenced to ~20x  
138 coverage each (for more details see M&M). This data was used to a) determine individual variability  
139 in allele frequencies and b) to validate the information content of the candidate SNP-set.

140 Using all individually resequenced individuals, we inferred the extent of genome wide linkage  
141 disequilibrium (LD;). The plot of LD  $r^2$  against the distance from the focal SNP showed that LD fell to  
142  $r^2 \sim 0.3$  within less than 120 bp, which means that genome positions such a distance apart are on  
143 average effectively unlinked (Fig. 3A). The PCA on SNP variation of the individually re-sequenced  
144 trees from the North population explained 12.3% of accumulated variation on the first two axes (Fig.  
145 3B). Trees from the same sampling site (within the North population) did not tend to cluster together  
146 (Fig. 3B).  $F_{ST}$  estimates among pools for non-overlapping 1 kb windows were virtually identical among  
147 healthy/damaged pools within region as compared to between regions (Suppl. Fig. 5). Trees within a  
148 phenotypic class were genomically not more similar than between classes (Suppl. Fig. 6, ANOSIM  $R =$   
149  $-0.008$ ,  $p = 0.76$ , 9,999 permutations).

150 Pool-GWAS analysis identified 106 SNPs significantly associated with the drought damage status  
151 using a Cochran-Mantel-Haenszel test on the two pairs of damaged and healthy pools after false  
152 discovery rate correction and a cut-off at  $1 \times 10^{-2}$  (Fig. 4A, Suppl. Fig. 7). Some of the 106 SNPs were  
153 in close physical proximity ( $<120$ bp) and thus probably linked. Taking this into account, 80  
154 independent genomic regions were associated with the drought damage status. None of the  
155 significantly differentiated SNP loci was mutually fixed; the observed allele frequency differences  
156 between healthy and damaged trees at associated loci ranged between 0.12 and 0.51 (Fig. 4B).

## 157 *Associated genes and gene function*

158 Of the 106 significant SNPs, 24 were found in 20 protein coding genes (Table 1). Forty-nine genes  
159 were the closest genes to the remaining 82 SNPs. For 61 of these genes, the best BLAST hit was with  
160 a tree, mainly from the Fagales genera *Quercus* and *Castanea* (Table 1, Suppl. Table 3). Among the 24  
161 SNPs in genes, we observed 13 non-synonymous changes. In eleven of these changes, the alternate  
162 allele was associated with the damaged phenotype and only in two cases with the healthy  
163 phenotype. Three of the non-synonymous substitutions resulted in a stop codon. Of the remaining  
164 ten, eight exchanges caused a major change in amino acid characteristics and thus probably in  
165 protein folding or function (Table 1). One gene, a PB1 domain-containing protein tyrosine kinase,  
166 contained four non-synonymous changes, suggesting that the allele version associated with the  
167 damaged phenotype lost its function (Tab. 1). From the 20 genes with significant SNPs, functional  
168 information could be obtained from the UniProt database for 14 (Suppl. Table 3). Of these, ten genes  
169 were associated in previous studies with either environmental stress response (two) or specifically  
170 with drought stress response (eight; Suppl. Table 3). Of the 49 predicted genes closest to the  
171 remaining significant SNPs (Tab. 1), 16 could be reliably annotated (Suppl. Table 2). Twelve had been  
172 directly related to drought in previous studies, while three were previously associated with other  
173 environmental stress responses (Suppl. Tab. 3).

## 174 *Genomic prediction*

175 We furthermore set out to determine how many SNPs were needed to successfully predict the  
176 drought susceptibility of individual trees, i.e. to develop a genotyping assay. All Pool-GWAS SNPs in  
177 addition to the top 20 individual re-sequencing SNPs were used to create a SNP combination to reach  
178 a genotyping success threshold of min. 90%. After excluding loci due to technical reasons and  
179 filtering for genotyping success, seventy loci proved to be suitable for reliable genotyping with a SNP  
180 assay. We genotyped only individuals sampled in 2019 that were not used to identify the SNPs in the  
181 first place plus paired individuals sampled in Aug 2020. On average, each of the 95 individuals was  
182 successfully genotyped at 67.7 loci (96.7%). We coded the genotypes as 0 for homozygous reference  
183 allele, 1 for heterozygous and 2 for the homozygous alternate allele, thus assuming a linear effect  
184 relationship. Figure 5 shows the genotypogram for the tested individuals.

185 Linear discriminant analysis (LDA) correctly predicted the observed phenotype from the genotype in  
186 91 of 92 cases (98.9%). Prediction success decreased to 65% when successively removing loci from  
187 the analysis (Suppl. Fig. 8). Nevertheless, ordering the individuals according to the LDA score of axis 1  
188 revealed no clear genotype pattern that distinguished healthy from damaged trees (Fig. 5). Observed  
189 heterozygosity at loci used in the SNP assay of individuals in the upper half of predictive values for a  
190 healthy phenotype was not significantly different from heterozygosity of the lower half (Suppl. Fig.  
191 9). Ordering the loci according to their squared loadings showed that loci's contribution to the  
192 genomic prediction differed substantially (Fig. 5). As expected, the histogram of LDA scores showed  
193 two peaks, corresponding to the two phenotypes (Suppl. Fig. 10).

194 To validate the results of the LDA prediction and to circumvent potential overfitting due to the small  
195 sample size, we also applied a non-parametric Machine Learning algorithm for feature selection and  
196 clustering that was especially designed for small sample sizes (Gerber2020, Horenko2020). The  
197 Method identified the 20 most-significant SNPs allowing to make an almost 85% correct classification  
198 that distinguished healthy from damaged trees (Suppl. Table 5).

## 199 **Discussion**

200 Over the last two decades, increasing drought periods caused severe damage to European forests  
201 (Schuldt *et al.* 2020b; Etzold *et al.* 2019; Pretzsch *et al.* 2013). Conifers seem to suffer the most, but  
202 also deciduous trees were strongly affected (Schuldt *et al.* 2020b). Weather data from our study area  
203 from 1950 onwards suggested that the climatic conditions for beech trees in the area investigated  
204 changed dramatically during this period. Roughly estimating the tree age from their trunk  
205 circumference (Bošela *et al.* 2014), more than a third of the trees were already in place at the  
206 beginning of this period. About 60% were recruited prior to the acceleration of temperature change  
207 from the 1980s onwards. As a result, trees in the mountainous regions of the study area today  
208 experience climatic conditions comparable to those experienced by low land trees in the 1950s,  
209 which in turn now experience a climate that used to be typical for regions much further South. Given  
210 the documented propensity of beech for local adaptation (Gárate-Escamilla *et al.* 2019; Pluess *et al.*  
211 2016; Aranda *et al.* 2015), including drought (Bolte *et al.* 2016), it is therefore conceivable that  
212 current conditions exceed the reaction norm of some previously locally well-adapted genotypes with  
213 detrimental consequences for their fitness. If the trend of an increasingly drier vegetation period  
214 persists, this will likely affect an even larger proportion of the currently growing beeches.

215 Evolutionary genomics will be indispensable to predict and manage the impact of global change on  
216 biodiversity (Waldvogel *et al.* 2020). As already shown for other partially managed (tree) species  
217 (Stocks *et al.* 2019), in particular pool-GWAS approaches (Endler *et al.* 2016) have proven to be useful  
218 in guiding conservation management.

219 Our strictly pairwise sampling design avoided many pitfalls of GWAS studies, arising, e.g., from  
220 cryptic population structure and shared ancestry (Hoban *et al.* 2016; Wellenreuther & Hansson  
221 2016). Despite presented evidence from this and other studies (Schuldt *et al.* 2020) that the observed  
222 crown damages in large parts of Central Europe used for phenotyping here are directly or indirectly  
223 due to the severe drought years 2018 and 2019, we must acknowledge that we have no direct  
224 physiological proof that the trees surveyed here indeed suffered from drought stress. In addition, the  
225 observed diagnostic symptoms are not specific to drought stress. Nevertheless, an unknown  
226 independent stressor would have needed to accidentally co-occur spatially and temporally with the  
227 drought. The phenotypical drought response of individual trees may also be influenced by  
228 microspatial variation (Carrière *et al.* 2020). In the present study, however, the mean distance  
229 between sampled paired trees of about 5 m assured that their roots systems largely overlapped.  
230 Thus, environmental variation in soil quality, rooting depth, water availability or other factors should  
231 have been minimal. Please note that any phenotypical misclassification due to such microspatial  
232 differences would have rather dissimulated the genotypic differences found in GWAS than enhanced  
233 them artificially. Also the lumping of similar phenotypes induced by different stressors is unlikely to  
234 have the same genomic basis and resulting in significant GWAS results.

235 As expected from previous studies (Rajendra *et al.* 2014), we found no population structure among  
236 the sampling sites. Applying relatively strict significance thresholds, we found systematic genomic  
237 differences between the healthy and damaged trees. In all cases, these differences were quantitative  
238 and not categorical, i.e. we found allele frequency changes but no fixed SNPs between phenotypes.  
239 Significant SNPs were mostly not clustered - we found on average 1.4 selected SNPs in a particular  
240 region. These findings were in line with the observed very short average LD in *F. sylvatica*, indicating  
241 that polymorphisms associated with the two phenotypes were likely old standing genetic variation  
242 (Harris & Nielsen 2013). Moreover, such SNPs are mostly detached from the background in which  
243 they arose and they are therefore often the actual causal variants. This observation is underlined by  
244 the high proportion of non-synonymous significant SNPs within genes, which in most cases caused  
245 substitution to amino acid with different properties or even premature termination. Such deviant  
246 variants with likely substantial functional or conformational changes in the resulting proteins may be  
247 selectively neutral or nearly neutral under ancestral benign conditions, but may become selectively  
248 relevant under changing conditions (Paabi & Rockman 2014). Interestingly, most of the allelic  
249 variants associated with a healthy phenotype were also the variants in the reference genome. This  
250 might be due to the choice of the *F. sylvatica* individual from which the reference genome was  
251 gained (Mishra *et al.* 2018). This more than 300-year-old individual is standing at a particularly dry  
252 site on a rocky outcrop on the rim of a scarp where precipitation swiftly runs off. Trees at such sites  
253 were likely selected for drought tolerance.

254 Even though the area sampled for this study was limited relative to the species distribution range, it  
255 comprised its core area. In addition, the climatic variation covered by the sampling sites for this study  
256 is representative for large parts of the species range (Baumbach *et al.* 2019). The relatively limited  
257 population structure over large parts of the species range (Magri *et al.* 2006) together with the  
258 propensity for long range gene-flow (Belmonte *et al.* 2008) suggested that the genomic variation  
259 responsible for drought tolerance identified here is widely distributed (Lander *et al.* 2021).  
260 Nevertheless, an assessment of the geographic distribution of the drought related genomic variants  
261 over the entire distribution range would yield general insight into the species-wide architecture of  
262 this important trait.

263 None of the genes found here was involved in a transcriptomic study on drought response in beech  
264 saplings (Müller *et al.* 2017). However, most of the reliably annotated genes with or close to SNP loci  
265 significantly associated with drought phenotypes had putative homologs in other plant species

266 previously shown to be involved in drought or different environmental stress response (for citations  
267 see Suppl. Table 2). This observation may be considered as post-hoc evidence that drought was  
268 indeed the most likely stressor causing the observed phenotypic responses. It remained unclear  
269 whether the remaining, not annotated genes had not yet been associated with drought before, or  
270 whether we were just unable to make this link them due to the lack of (ecological) annotation and  
271 standardised reporting. (Waldvogel *et al.* 2021) The involvement of in total 67 genes together with  
272 the relatively flat effect size distribution suggested that drought resistance in *F. sylvatica* is a  
273 moderately polygenic trait, which should respond well to artificial breeding attempts and natural  
274 selection. However, given the relatively strict threshold criteria, it is likely that more yet undetected  
275 loci contribute to the respective phenotypes. The low LD in beech predicts that an adaptation to  
276 drought will not compromise genome-wide genetic diversity and thus adaptation potential to other  
277 stressors. We achieved a high level of accuracy using genomic data to predict the drought phenotype  
278 from individuals not used to identify drought associated SNP loci. However, due to the small sample  
279 size, LDA might have resulted in overfitting (Hawkins 2004). We therefore also used a non-parametric  
280 machine learning algorithm that has been shown to produce more robust results, especially for small  
281 sample sizes (Horenko 2020). Both analyses confirmed that we mainly identified alleles widespread  
282 throughout the sampled range and not locally specific. Besides, we confirmed a considerable level of  
283 genetic variation in the sampled regions. The observation that trees with the highest predictive  
284 values showed no loss of heterozygosity indicated that there is still adaptive potential for drought  
285 adaptation in the species (Gienapp *et al.* 2017). With the SNP assay, we therefore created a tool that  
286 can i) support the choice of seed trees for reforestations, ii) provide decision guidance for selective  
287 logging and iii) monitor, whether natural selection on this quantitative trait is already acting in the  
288 species. The current study can also serve as a starting point for molecular and physiological research  
289 on how the identified loci or variants may, alone or in concert, confer resilience or tolerance to a  
290 range of drought stress symptoms.

## 291 **Material and Methods**

### 292 *Sampling and phenotyping*

293 In August/early September 2019, we sampled leaf tissue of 402 *Fagus sylvatica* trees from 32  
294 locations in Hessen/Germany (set 1, Figure 1), of which 300 were used for the (pool)GWAS analysis.  
295 Forty three, plus additional 53 trees which were sampled in n August 2020, additional 52 trees from  
296 four sites were sampled (set 2, Figure 1) made up the confirmation set. The coordinates and  
297 characteristics of each site can be found in Suppl. Table 1. The sampling was performed in a strictly  
298 pairwise design. The pairs consisted of one tree with heavy drought damage of the crown (lost or  
299 rolled up, dried leaves) and one with an unaffected crown, respectively. This categorisation into least  
300 and most damaged trees was taken compared to the other trees in the respective forest patch. The  
301 pairs were *a priori* chosen such that the two trees were i) mutually the closest neighbours with  
302 contrasting damage status (i.e. no other tree in the direct sight-line), ii) free from apparent  
303 mechanical damage, fungal infestations or other signs of illness, similar iii) in tree height, iv) trunk  
304 circumference, v) light availability, and vi) canopy closure. In addition, each pair was situated at least  
305 30 m from the closest forest edge. For each tree of the chosen pairs, we recorded the exact position,  
306 distance to the pair member and the estimated tree height (in 1 m increments), measured the trunk  
307 circumference at 150 cm height above the ground (in 10 cm increments), and estimated the leaf loss  
308 of the crown and the proportion of dried leaves (in 5% increments). We also recorded the estimated  
309 distance (in 1 m resolution) and the specific identity of the two closest neighbour trees for each pair  
310 member and calculated a competition index  $C$  as follows:  $C = S_1/D_1 + S_2/D_2$ , where  $S_1$  and  $S_2$  are the  
311 trunk diameter at 150 cm and  $D_1$  and  $D_2$  the distances of the nearest and second nearest neighbour

312 tree of the same size or larger than the focal tree. Photographs from the crown and the trunk were  
313 taken from the trees sampled in 2019.

314 From each tree, we sampled 5 to 10 fully developed leaves from low branches. The leaves sampled  
315 from each tree were placed in paper bags. After returning from the field, they were dried at 50°C for  
316 30-90 min and then kept on salt until they could be stored at -80°C.

#### 317 *Climate and remote sensing data*

318 Monthly daily mean minimum and maximum temperature values and precipitation data were  
319 obtained for the 1 x 1 km grid cells harbouring the sampling sites for the period between 1950 and  
320 2019. Data on the accumulated potential evapotranspiration during the growth season was obtained  
321 for the same grid cells. The data is publicly available from  
322 [https://opendata.dwd.de/climate\\_environment/CDC/grids\\_germany/monthly/](https://opendata.dwd.de/climate_environment/CDC/grids_germany/monthly/).

323 Leaf area index (LAI) data for the above grid cells was obtained from Copernicus remote sensing  
324 ([www.copernicus.eu](http://www.copernicus.eu)) for the period 2014-2019, considering only the month of August. To see  
325 whether drought conditions influenced leaf coverage of the woods at the sampling sites, we  
326 calculated the relative annual deviation of LAI from the 2014 value. We correlated it to the relative  
327 deviation of the cumulated potential evapotranspiration over the growth season from 2014. The year  
328 2014 was used as a baseline, because of the significant drought increase since then (Büntgen *et al.*  
329 2021). Please note that the absolute level of LAI depends on the wood coverage, vegetation density  
330 and species composition of each plot. Changes in LAI are thus not exclusively due to drought  
331 damages in beech.

#### 332 *DNA extraction, construction of GWAS pools and sequencing*

333 DNA was extracted from 12.5 mm<sup>2</sup> of a single leaf from each tree following the NucleoMag Plant Kit  
334 (Macherey Nagel, Düren, Germany) protocol. We set up four DNA pools for poolGWAS by pooling  
335 equal amounts of DNA from each individual: damaged individuals from the Southern part (dSouth),  
336 healthy individuals from the South (hSouth), damaged North (dNorth) and healthy North (hNorth).  
337 The Southern pools consisted of 100 individuals each, the Northern pools of 50 individuals each. The  
338 pools were sent to Novogene (Cambridge, UK) for library construction and 150bp paired end  
339 sequencing with 350bp insert size with 25Gb data for the northern and 38Gb data for the southern  
340 samples. The 100 individuals used to construct the Northern pools were also individually re-  
341 sequenced. The exact composition of the genomic pools can be found in Supplemental Table 1. All  
342 sequence information can be found on the European Nucleotide Archive (ENA) under project  
343 accession number *PRJEB24056*.

#### 344 *Reference genome improvement*

345 We used an improved version of the recently published reference genome for the European beech  
346 (Mishra *et al.* 2018). Contiguity was improved to chromosome level using Hi-C reads with the help of  
347 the allhic software after excluding the probable organelle backbones from the earlier assembly that  
348 was generated from the Illumina-corrected PacBio reads using Canu assembler (Mishra *et al.* 2021)  
349 Access. No. PRJNA450822.

#### 350 *Mapping and variant calling*

351 Reads of pools and individual resequencing were trimmed using the wrapper tool autotrim v0.6.1  
352 (Waldvogel *et al.* 2018) that integrates trimmomatic (Bolger *et al.* 2014) for trimming and fastQC  
353 (Andrews 2010) for quality control. The trimmed reads were then mapped on the latest chromosome  
354 level build of the *F. sylvatica* genome using the BWA mem algorithm v.0.7.17 (Li & Durbin 2009). Low

355 quality reads were subsequently filtered and SNPs were initially called using samtools v.1.10 (Li *et al.*  
356 2009). A principal component analysis (PCA) was conducted on unlinked single nucleotide  
357 polymorphisms (SNPs) using the R package Factoextra v.1.0.7 (Kassambara & Mundt 2017).

#### 358 *Pool GWAS and PLINK*

359 The PoPoolation pipeline 2\_2012 (Kofler *et al.* 2011a; Kofler *et al.* 2011b) was used to call SNPs and  
360 remove indels from the four pools. Allele frequencies for all SNPs with a coverage between 15x and  
361 100x with a minimum allele count of three were estimated with the R library PoolSeq v. 0.35 (Taus *et*  
362 *al.* 2017).

363 The statistical test to detect significant allele frequency differences among damaged and healthy  
364 trees was the Cochran-Mantel-Haenszel test. With this test, a 2x2 table was created for each variable  
365 position and region with two phenotypes (healthy and damaged). The read counts of each allele for  
366 each phenotype were treated as the dependent variables. We controlled for false discovery rate  
367 using the Benjamini-Hochberg correction R package *p.adjust*.

368 For the individual resequencing data we followed the GATK-pipeline 4.1.3.0 (DePristo *et al.* 2011). In  
369 short, Picard tools v.2.20.8 was used to mark duplicates. GVCF files were created with  
370 HaplotypeCaller and genotyped with GEnotypeGVCFs. Since we did not have a standard SNP set we  
371 hard filtered SNPs with VariantFiltration QD<2.0, MQ<50.0, MQRankSum<12.5,  
372 ReadPosRankSum<8.0, FS>80.0, SOR>4.0 and QUAL<10.0. This conservative SNP-set was used for  
373 base recalibration before running the HaplotypeCaller pipeline a second round. Finally, the  
374 genotyped vcf-files were filtered using vcftools with --maf 0.03 --max-missing 0.9 --minQ 25 --min-  
375 meanDP 10 --max-meanDP 50 --minDP 10 --maxDP 50. The detailed pipeline can be found in Suppl.  
376 Info 2.

377 To conduct the GWAS association on the above generated SNP set with phenotypes being either  
378 damaged or healthy and to generate a principal component analysis on the SNP positions of the  
379 individually resequenced trees, we used PLINK 1.9 (Purcell *et al.* 2007). The detailed workflow can be  
380 found in Suppl. Info 2. We calculated a non-parametric ANOSIM on an inter-individual Euclidean  
381 distance matrix based on the first ten principal components to infer whether the trees within  
382 phenotype groups are overall genetically more similar than within groups (9,999 permutations;  
383 (Hammer *et al.* 2001).

#### 384 *Inference of Linkage Disequilibrium*

385 The expected length of segregating haplotypes in a species depends on the recombination rate and  
386 their age. The former can be approximated by an estimate of linkage disequilibrium (LD). To  
387 determine LD decay based on individually re-sequenced data we used the software LDkit v 1.0.0  
388 (Tang *et al.* 2020), in 1kb and 100kb windows.

#### 389 *Identification substitution type and gene function*

390 We inferred whether significantly differentiated SNPs within genes lead to a (non-) synonymous  
391 amino acid substitution using tbg-tools v0.2 (<https://github.com/Croxa/tbg-tools>) (Schoennenbeck *et*  
392 *al.* 2021). The protein sequences of the identified genes were used in a blastp search against all non-  
393 redundant GenBank CDS translations, PDB, SwissProt, PIR, PRF to infer potential gene functions. Only  
394 the best BLAST-hits were considered.

#### 395 *Selection of SNP loci for SNPtype™ assay design*

396 For the design of SNPtype™ assays we used the web-based D3 assay design tool (Fluidigm corp.). We  
397 aimed in first preference for the most significant SNPs of each genomic region identified by Pool-  
398 GWAS (80 loci). If this was technically impossible and the region harboured more than a single  
399 significant SNP, we opted for the second most significant SNP and so forth. This resulted finally in 76  
400 suitable loci. The remaining 20 loci were recruited from the 20 most significant SNPs of the PLINK  
401 analysis that were not scored in the Pool-GWAS.

#### 402 *SNP genotyping procedure*

403 For validation of drought susceptibility associated SNPs, we conducted SNP genotyping on 96.96  
404 Dynamic Arrays (Fluidigm) with integrated fluidic circuits (Wang *et al.* 2009, 2009). (N=96) to validate  
405 the effectiveness of the identified SNPs in discriminating healthy from damaged trees. Prior to  
406 genotyping PCR, DNA extracts were normalised to approximately 5-10 ng/μl. They underwent a pre-  
407 amplification PCR (Specific Target Amplification, STA) according to the manufacturer's protocol to  
408 enrich target loci. PCR products were diluted 1:10 with DNA suspension buffer (TEKnova, PN T0221)  
409 before further use. Genotyping was performed according to the recommendations of manufacturer.  
410 Four additional PCR cycles were added to accommodate for samples of lower quality or including  
411 inhibitors (von Thaden, 2020). Fluorescent data were measured using the EP1 (Fluidigm) and  
412 analysed with the SNP Genotyping Analysis Software version 4.1.2 (Fluidigm). The automated scoring  
413 of the scatter plots was checked visually and if applicable, manually corrected.

#### 414 *Genomic prediction*

415 To predict drought susceptibility from genotype data, we used a linear discriminant analysis (LDA) on  
416 92 genotypes scored with the Fluidigm assay at 70 loci. Genotypes homozygous for the reference  
417 allele were scored as 0, heterozygous as 1 and homozygous alternate alleles as 2. We used the LDA  
418 option implemented in PAST v. 4.05. (Hammer, 2001).

419 We also used a non-parametric entropy-based Scalable Probabilistic Analysis framework (eSPA). This  
420 method allows simultaneous solution of feature selection and clustering problems, meaning that  
421 does not rely on a particular choice of user-defined parameters and has been shown to produce  
422 more robust results, especially for small sample sizes (Gerber2020, Horenko2020). Following the  
423 suggestion of the user manual, eSPA analysis was run 100 times with independent cross-validations  
424 of the Area Under the Curve (AUC) on the validation data.

## 425 **Acknowledgments**

426 There was no external funding for this study. We thank the SBIK-F Data- and Modelling Centre (Aidin  
427 Niamir) for support in obtaining climate and remote sensing data. CC acknowledges support from the  
428 Emergent AI Center funded by the Carl-Zeiss-Stiftung.

## 429 **Additional information**

### 430 *Author contributions*

431 Markus Pfenninger: Conceptualisation, sampling, data analysis, writing – original draft. Friederike  
432 Reuss, Angelika Kiebler: Laboratory support. Philipp Schönnenfeld, Cosima Caliendo, Susanne Gerber:  
433 Bioinformatics support. Bernardino Cocchiararo: SNP-assay. Sabrina Reuter, Nico Blüthgen, Karsten  
434 Mody: Conceptualisation, sampling, data analysis. Miklós Bálint: Conceptualisation, data analysis.  
435 Bagdevi Mishra, Marco Thines: Bioinformatics support. Barbara Feldmeyer: Conceptualisation,  
436 labwork, data analysis. All authors read and approved the final version of the manuscript.

### 437 *Author ORCIDs*

438 Markus Pfenninger 0000-0002-1547-7245

439 Susanne Gerber 0000-0001-9513-0729

440 Berardino Cocchiararo 0000-0002-8427-7861

441 Karsten Mody 0000-0002-9251-0803

442 Marco Thines 0000-0002-9251-0803

443 Barbara Feldmeyer 0000-0002-0413-7245

444 *Data availability*

445 Sequencing data have been deposited at ENA under project code PRJEB41889.

446 The genome assembly including the annotation is available under the Access. No. PRJNA450822.

## 447 **References**

448 Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Babraham  
449 Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

450 Aranda I, Cano FJ, Gascó A, *et al.* (2015) Variation in photosynthetic performance and hydraulic  
451 architecture across European beech (*Fagus sylvatica* L.) populations supports the case for  
452 local adaptation to water stress. *Tree Physiology* **35**, 34-46.

453 Bastide H, Betancourt A, Nolte V, *et al.* (2013) A genome-wide, fine-scale map of natural  
454 pigmentation variation in *Drosophila melanogaster*. *PLoS Genetics* **9**, e1003534.

455 Baumbach L, Niamir A, Hickler T, Yousefpour R (2019) Regional adaptation of European beech (*Fagus*  
456 *sylvatica*) to drought in Central European conditions considering environmental suitability  
457 and economic implications. *Regional Environmental Change* **19**, 1159-1174.

458 Belmonte J, Alarcón M, Avila A, Scialabba E, Pino D (2008) Long-range transport of beech (*Fagus*  
459 *sylvatica* L.) pollen to Catalonia (north-eastern Spain). *International Journal Of*  
460 *Biometeorology* **52**, 675-687.

461 Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data.  
462 *Bioinformatics* **30**, 2114-2120.

463 Bolte A, Czajkowski T, Coccozza C, *et al.* (2016) Desiccation and mortality dynamics in seedlings of  
464 different European beech (*Fagus sylvatica* L.) populations under extreme drought conditions.  
465 *Frontiers in plant science* **7**, 751.

466 Bošela M, Sedmák R, Marušák R, *et al.* (2014) Evaluating similarity of radial increments around tree  
467 stem circumference of European beech and Norway spruce from Central Europe.  
468 *Geochronometria* **41**, 136-146.

469 Bressemer U (2008) Komplexe Erkrankungen an Buche. Complex diseases in beech. *Ergebnisse*  
470 *angewandter Forschung zur Buche* **3**, 87.

471 Brunet J, Fritz Ö, Richnau G (2010) Biodiversity in European beech forests-a review with  
472 recommendations for sustainable forest management. *Ecological Bulletins*, 77-94.

473 Büntgen U, Urban O, Krusic PJ, *et al.* (2021) Recent European drought extremes beyond Common Era  
474 background variability. *Nature Geoscience*, 1-7.

475 Carrière SD, Ruffault J, Cakpo CB, *et al.* (2020) Intra-specific variability in deep water extraction  
476 between trees growing on a Mediterranean karst. *Journal of Hydrology* **590**, 125428.

477 Choat B, Brodribb TJ, Brodersen CR, *et al.* (2018) Triggers of tree mortality under drought. *Nature*  
478 **558**, 531-539.

479 Christensen JH, Hewitson B, Busuioc A, *et al.* (2007) Regional climate projections. In: *Climate Change,*  
480 *2007: The Physical Science Basis. Contribution of Working group I to the Fourth Assessment*  
481 *Report of the Intergovernmental Panel on Climate Change, University Press, Cambridge,*  
482 *Chapter 11*, pp. 847-940.

483 Coccozza C, De Miguel M, Pšidová E, *et al.* (2016) Variation in ecophysiological traits and drought  
484 tolerance of beech (*Fagus sylvatica* L.) seedlings from different populations. *Frontiers in Plant*  
485 *Science* **7**, 886.

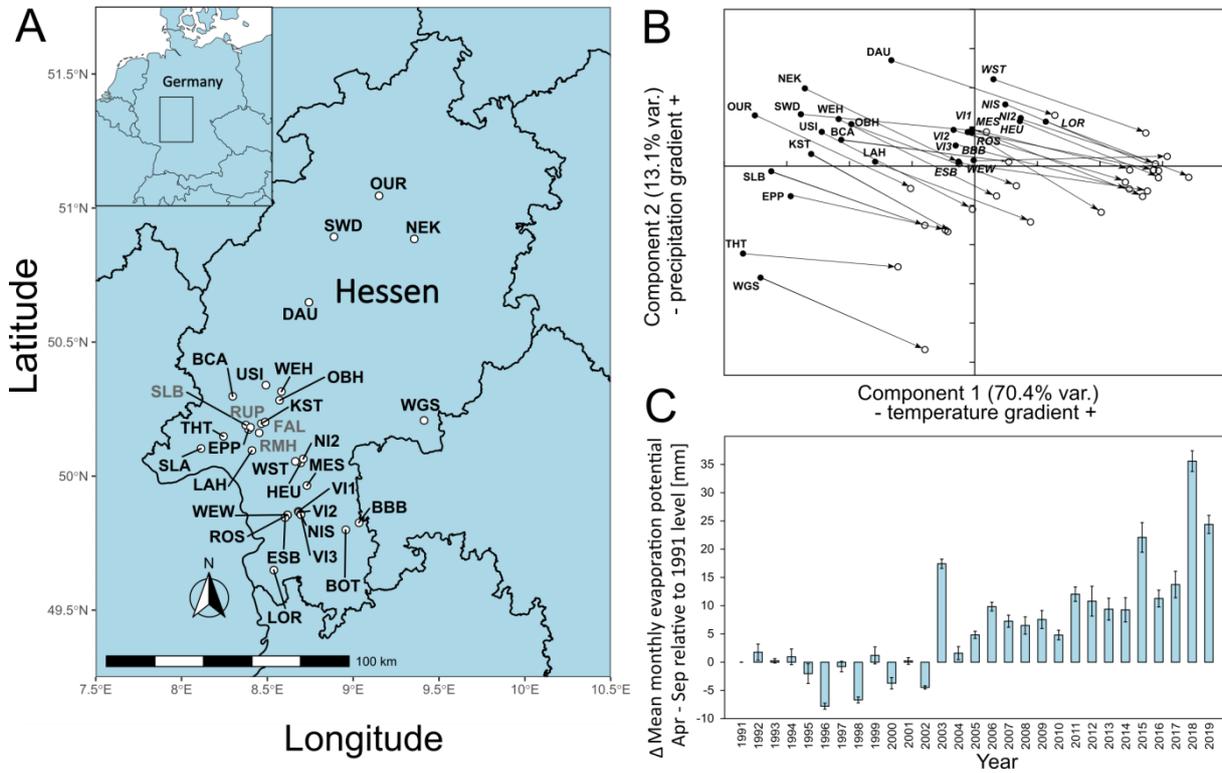
- 486 Czajkowski T, Bolte A (2006) Frosttoleranz deutscher und polnischer Herkünfte der Buche (*Fagus*  
487 *sylvatica* L.) und ihre Beeinflussung durch Trockenheit. *Archiv für Forstwesen und*  
488 *Landschaftsökologie* **40**, 119-126.
- 489 DePristo MA, Banks E, Poplin R, *et al.* (2011) A framework for variation discovery and genotyping  
490 using next-generation DNA sequencing data. *Nature Genetics* **43**, 491.
- 491 Dorow WH, Blick T, Kopelke J-P (2010) Zoologische Forschung in hessischen Naturwaldreservaten–  
492 Exemplarische Ergebnisse und Perspektiven. *Forstarchiv* **81**, 61-68.
- 493 Elsasser P, Meyerhoff J, Weller P (2016) An updated bibliography and database on forest ecosystem  
494 service valuation studies in Austria, Germany and Switzerland. Thünen Working Paper.
- 495 Endler L, Betancourt AJ, Nolte V, Schlötterer C (2016) Reconciling differences in pool-GWAS between  
496 populations: a case study of female abdominal pigmentation in *Drosophila melanogaster*.  
497 *Genetics* **202**, 843-855.
- 498 Etzold S, Ziemińska K, Rohner B, *et al.* (2019) One century of forest monitoring data in Switzerland  
499 reveals species- and site-specific trends of climate-induced tree mortality. *Frontiers in Plant*  
500 *Science* **10**, 307.
- 501 Gárate-Escamilla H, Hampe A, Vizcaíno-Palomar N, Robson TM, Benito Garzón M (2019) Range-wide  
502 variation in local adaptation and phenotypic plasticity of fitness-related traits in *Fagus*  
503 *sylvatica* and their implications under climate change. *Global Ecology and Biogeography* **28**,  
504 1336-1350.
- 505 Gerber S, Pospisil L, Navandar M, Horenko I (2020). Low-cost scalable discretization, prediction, and  
506 feature selection for complex systems. *Science Advances*, **6**, eaaw0961.
- 507 Gienapp P, Fior S, Guillaume F, *et al.* (2017) Genomic quantitative genetics to study evolution in the  
508 wild. *Trends in Ecology & Evolution* **32**, 897-908.
- 509 Harris K, Nielsen R (2013) Inferring demographic history from a spectrum of shared haplotype  
510 lengths. *PLoS Genet* **9**, e1003521.
- 511 Harter DE, Nagy L, Backhaus S, *et al.* (2015) A comparison of genetic diversity and phenotypic  
512 plasticity among European beech (*Fagus sylvatica* L.) populations from Bulgaria and Germany  
513 under drought and temperature manipulation. *International Journal of Plant Sciences* **176**,  
514 232-244.
- 515 Hawkins DM (2004) The problem of overfitting. *Journal of chemical information and computer*  
516 *sciences* **44**, 1-12.
- 517 Hoban S, Kelley JL, Lotterhos KE, *et al.* (2016) Finding the genomic basis of local adaptation: pitfalls,  
518 practical solutions, and future directions. *The American Naturalist* **188**, 379-397.
- 519 Horenko I (2020) On a Scalable Entropic Breaching of the Overfitting Barrier for Small Data Problems  
520 in Machine Learning. *Neural Computation* **32**, 1563-1579.
- 521 Kassambara A, Mundt F (2017) Factoextra: extract and visualize the results of multivariate data  
522 analyses. *R package version 1*, 337-354.
- 523 Kofler R, Orozco-terWengel P, De Maio N, *et al.* (2011a) PoPoolation: A Toolbox for Population  
524 Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. *PLoS One* **6**,  
525 e15925.
- 526 Kofler R, Pandey RV, Schlötterer C (2011b) PoPoolation2: identifying differentiation between  
527 populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **27**, 3435-  
528 3436.
- 529 Kreyling J, Thiel D, Nagy L, *et al.* (2012) Late frost sensitivity of juvenile *Fagus sylvatica* L. differs  
530 between southern Germany and Bulgaria and depends on preceding air temperature.  
531 *European Journal of Forest Research* **131**, 717-725.
- 532 Lander TA, Klein EK, Roig A, Oddou-Muratorio S (2021) Weak founder effects but significant spatial  
533 genetic imprint of recent contraction and expansion of European beech populations.  
534 *Heredity* **126**, 491-504.
- 535 Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform.  
536 *Bioinformatics* **25**, 1754-1760.

- 537 Li H, Handsaker B, Wysoker A, *et al.* (2009) The Sequence alignment/map (SAM) format and  
538 SAMtools. . *Bioinformatics* **25**, 2078-2079.
- 539 Magri D, Vendramin GG, Comps B, *et al.* (2006) A new scenario for the Quaternary history of  
540 European beech populations: palaeobotanical evidence and genetic consequences. *New*  
541 *Phytologist* **171**, 199-221.
- 542 Mishra B, Gupta DK, Pfenninger M, *et al.* (2018) A reference genome of the European beech (*Fagus*  
543 *sylvatica* L.). *Gigascience* **7**, giy063.
- 544 Mishra B, Ulaszewski B, Meger J, *et al.* (2021) A chromosome-level genome assembly of the  
545 European Beech (*Fagus sylvatica*) reveals anomalies for organelle DNA integration, repeat  
546 content and distribution of SNPs. *bioRxiv* <https://doi.org/10.1101/2021.03.22.436437>.
- 547 Müller M, Seifert S, Lübke T, Leuschner C, Finkeldey R (2017) De novo transcriptome assembly and  
548 analysis of differential gene expression in response to drought in European beech. *PLoS One*  
549 **12**, e0184167.
- 550 Paaby A, Rockman M (2014) Cryptic genetic variation: evolution's hidden substrate. *Nature Reviews*  
551 *Genetics* **15**, 247–258.
- 552 Paar U, Dammann I (2019) Waldzustandsbericht Hessen 2019 (ed. Hessisches Ministerium für  
553 Umwelt K, Landwirtschaft und Verbraucherschutz), Wiesbaden.
- 554 Pluess AR, Frank A, Heiri C, *et al.* (2016) Genome–environment association study suggests local  
555 adaptation to climate at the regional scale in *Fagus sylvatica*. *New Phytologist* **210**, 589-601.
- 556 Pretzsch H, Schütze G, Uhl E (2013) Resistance of European tree species to drought stress in mixed  
557 versus pure forests: evidence of stress release by inter-specific facilitation. *Plant Biology* **15**,  
558 483-495.
- 559 Purcell S, Neale B, Todd-Brown K, *et al.* (2007) PLINK: a tool set for whole-genome association and  
560 population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559-575.
- 561 Rajendra K, Seifert S, Prinz K, Gailing O, Finkeldey R (2014) Subtle human impacts on neutral genetic  
562 diversity and spatial patterns of genetic variation in European beech (*Fagus sylvatica*). *Forest*  
563 *Ecology and Management* **319**, 138-149.
- 564 Schoennenbeck P, Schell T, Gerber S, Pfenninger M (2021) tbg-a new file format for genomic data.  
565 *bioRxiv* <https://doi.org/10.1101/2021.03.15.435393>.
- 566 Schuldt B, Buras A, Arend M, *et al.* (2020a) A first assessment of the impact of the extreme 2018  
567 summer drought on Central European forests. *Basic and Applied Ecology* **45**, 86-103.
- 568 Schuldt B, Buras A, Arend M, *et al.* (2020b) A first assessment of the impact of the extreme 2018  
569 summer drought on Central European forests. *Basic and Applied Ecology*.
- 570 Stocks JJ, Metheringham CL, Plumb WJ, *et al.* (2019) Genomic basis of European ash tree resistance  
571 to ash dieback fungus. *Nature Ecology & Evolution* **3**, 1686-1696.
- 572 Sutmöller J, Spellmann H, Fiebigler C, Albert M (2008) Der Klimawandel und seine Auswirkungen auf  
573 die Buchenwälder in Deutschland The effects of climate change on beech forests in Germany.  
574 *Ergeb. Angew. Forsch. Buche* **3**, 135.
- 575 Tang Y, Li Z, Wang C, *et al.* (2020) LDkit: a parallel computing toolkit for linkage disequilibrium  
576 analysis. *BMC Bioinformatics* **21**, 1-8.
- 577 Taus T, Futschik A, Schlötterer C (2017) Quantifying selection with pool-seq time series data.  
578 *Molecular Biology and Evolution* **34**, 3023-3034.
- 579 Thünen\_Institute (2020) Clusterstatistik Forst & Holz.
- 580 Trenberth KE, Dai A, Van Der Schrier G, *et al.* (2014) Global warming and changes in drought. *Nature*  
581 *Climate Change* **4**, 17-22.
- 582 Waldvogel AM, Feldmeyer B, Rolshausen G, *et al.* (2020) Evolutionary genomics can improve  
583 prediction of species' responses to climate change. *Evolution Letters* **4**, 4-18.
- 584 Waldvogel A-M, Schreiber D, Pfenninger M, Feldmeyer B (2021) Climate Change Genomics Calls for  
585 Standardized Data Reporting. *Coping with Climate Change: A Genomic Perspective on*  
586 *Thermal Adaptation*.
- 587 Waldvogel AM, Wieser A, Schell T, *et al.* (2018) The genomic footprint of climate adaptation in  
588 *Chironomus riparius*. *Molecular Ecology* **27**, 1439-1456.

589 Wang J, Lin M, Crenshaw A, *et al.* (2009) High-throughput single nucleotide polymorphism  
 590 genotyping using nanofluidic Dynamic Arrays. *BMC Genomics* **10**, 561.  
 591 Wellenreuther M, Hansson B (2016) Detecting polygenic evolution: problems, pitfalls, and promises.  
 592 *Trends in Genetics* **32**, 155-164.

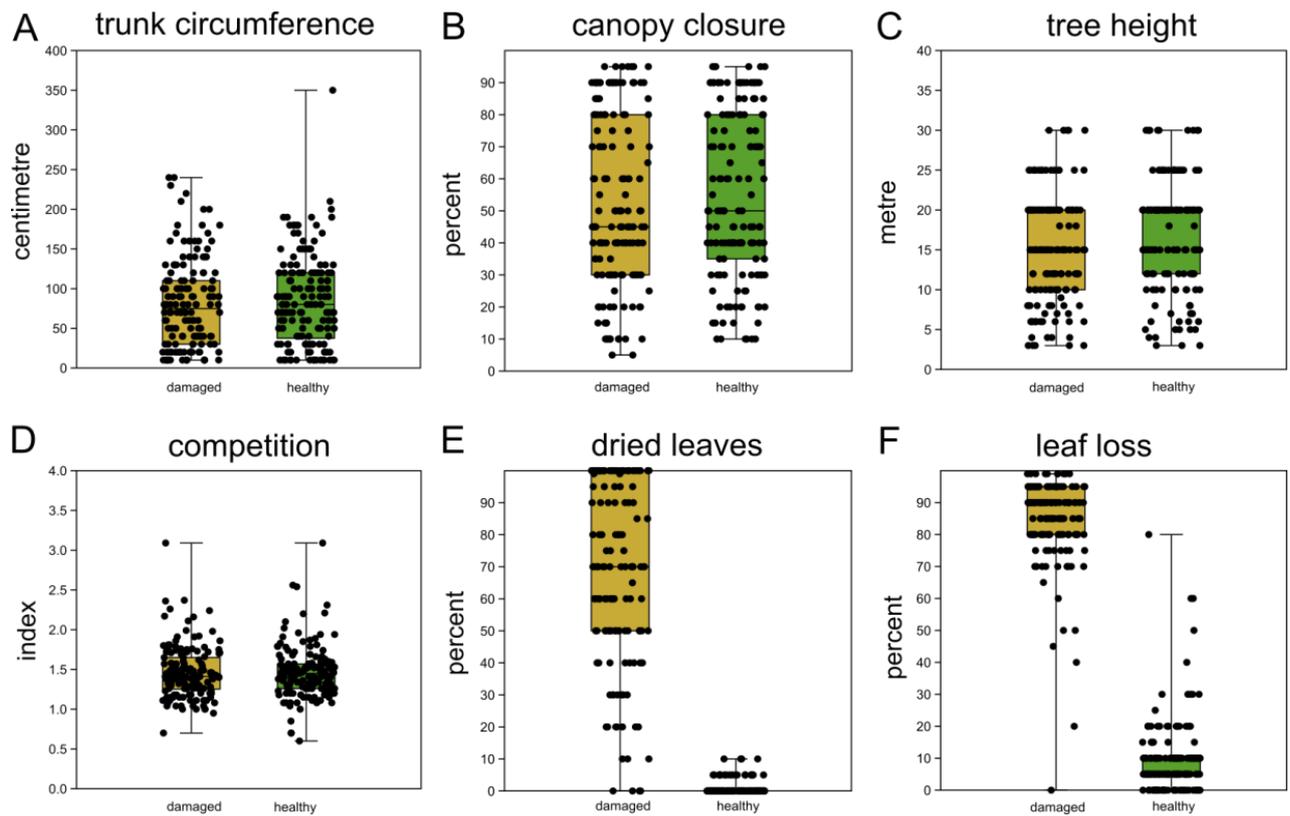
593

594 **Figures**



595

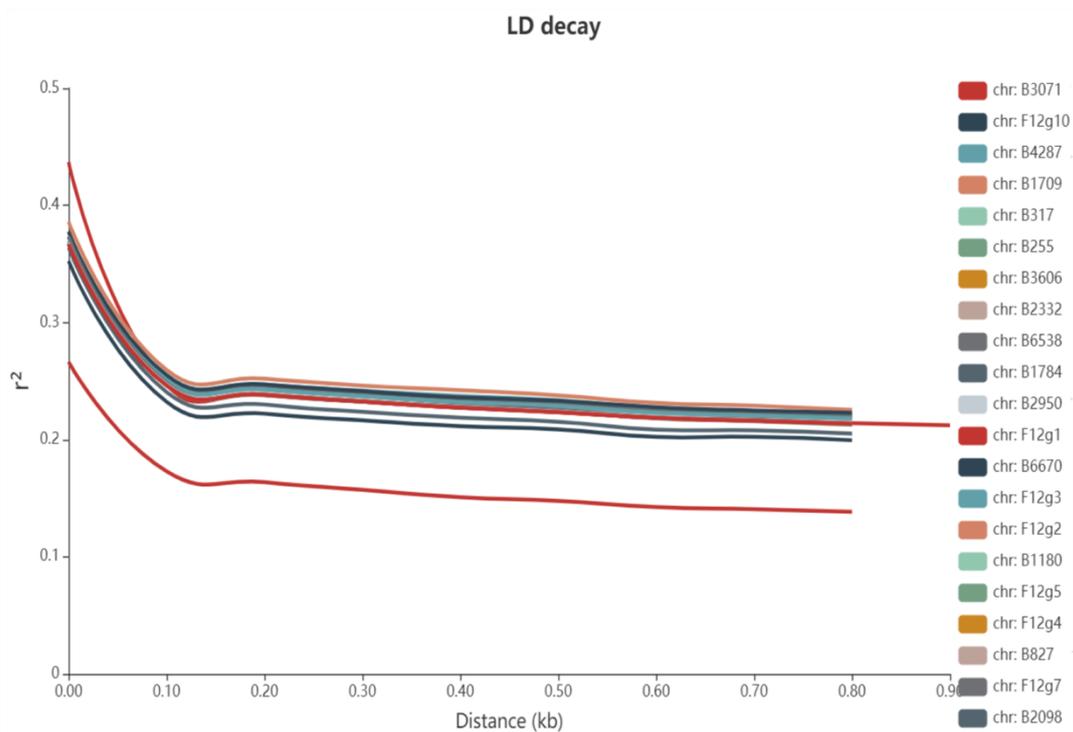
596 Figure 1. A) Locations of sampling sites in Hessen, Germany. For abbreviations see Suppl. Table 1. The  
 597 sites where confirmation individuals were sampled are designated in grey. B) Principal Component  
 598 Analysis of monthly climate data 1950-2019, C) Development of main growth period drought  
 599 indicator from 1991-2019. Shown is the difference mean monthly evaporation potential in mm from  
 600 April to September relative to the 1991 level. Climate and drought data obtained from  
 601 [https://opendata.dwd.de/climate\\_environment/CDC/grids\\_germany/monthly/](https://opendata.dwd.de/climate_environment/CDC/grids_germany/monthly/).



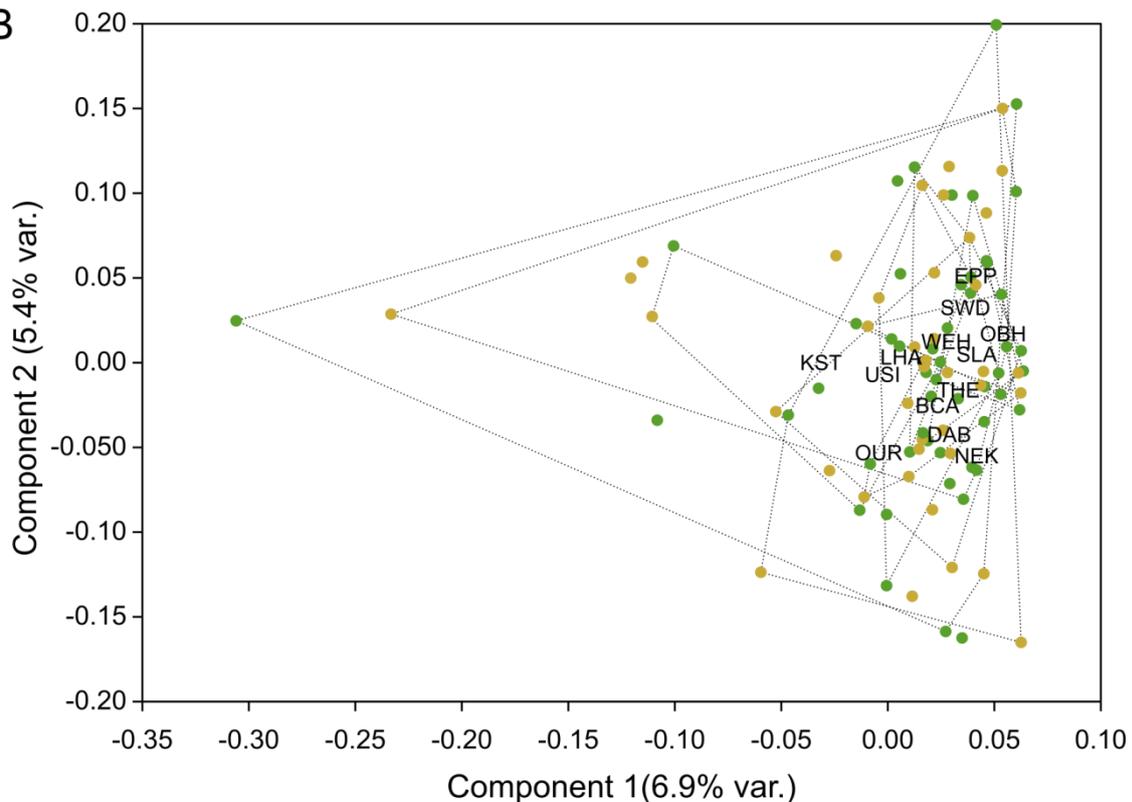
602

603 Figure 2. Comparison of sampled beech pairs. A) trunk circumference, B) canopy closure, C) tree  
604 height, D) competition index, E) dried leaves and F) leaf loss. Box-plots with indicated means, the  
605 boxes represent one standard deviation, the whiskers are the 95% confidence intervals. Damaged  
606 trees in ochre, healthy trees in green. Except for E and F, the difference of means among damaged  
607 and healthy trees is insignificant between the groups.

A



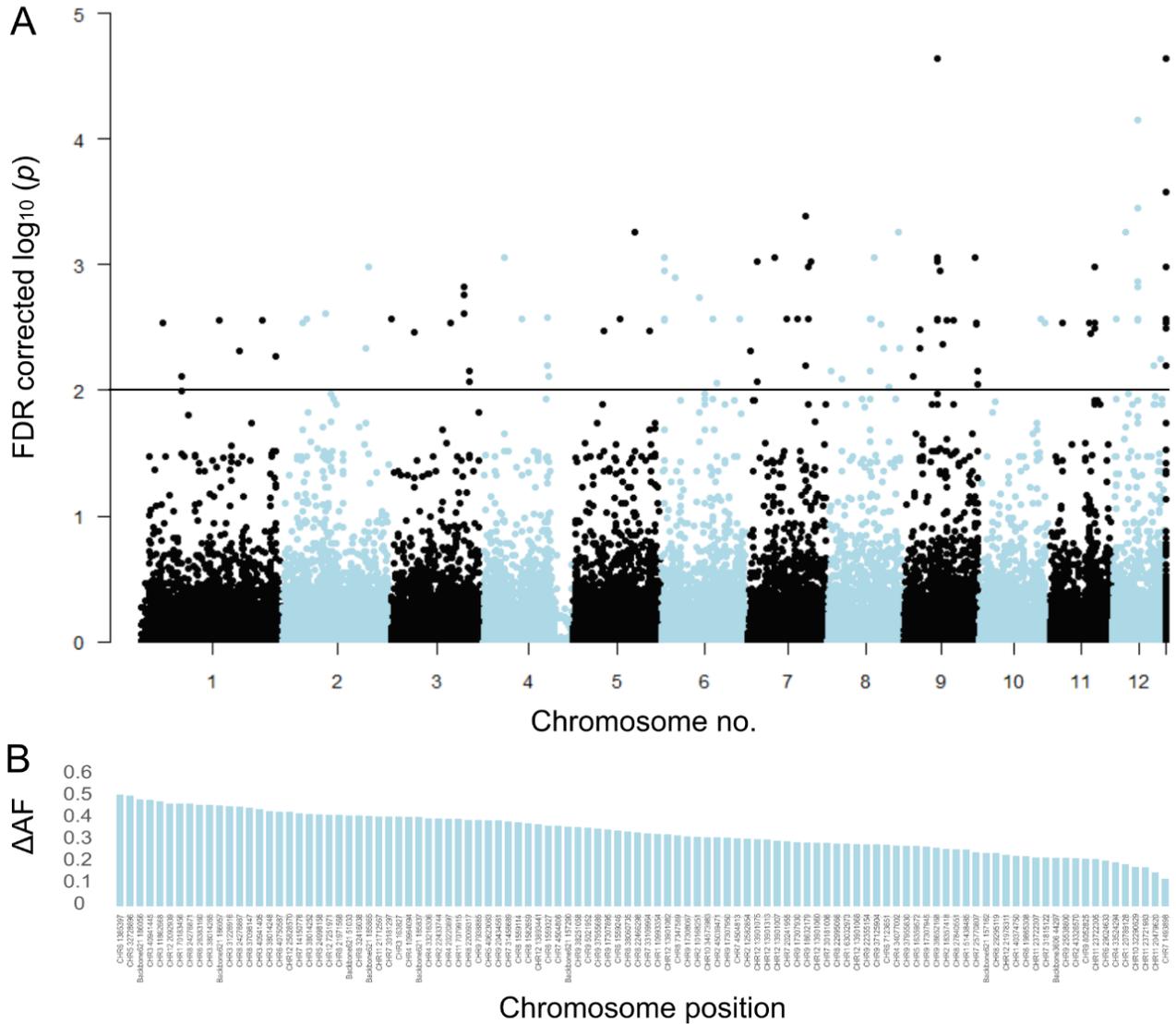
B



608

609 Figure 3. Genome wide linkage disequilibrium and principal component analysis on genome-wide  
610 SNP data. A) Decay of genome wide linkage disequilibrium (LD), measured as  $r^2$  on allele frequencies  
611 gained from individual resequencing, with distance from focal SNP in base pairs. B) Plot of the first  
612 two principal component axes of LD pruned SNP data from individually sequenced beech individuals

613 from the North population. Healthy trees are indicated by a green dot, damaged ones by ochre.  
614 Individuals sampled from the same site are grouped by convex hulls, limited with dotted lines.  
615



616

617 Figure 4. Significantly associated drought phenotype associated SNP marker. A) Manhattan plot of  
 618 false discovery rate (FDR) corrected  $-\log_{10}$  probability values from CMH test. The black horizontal  
 619 line indicates the chosen significance threshold. SNPs on different chromosomes alternate in colour  
 620 (black and blue). B) Mean allele frequency difference at significantly associated SNP loci between  
 621 healthy and damaged phenotypes. The loci are ordered according to amount of change.

622

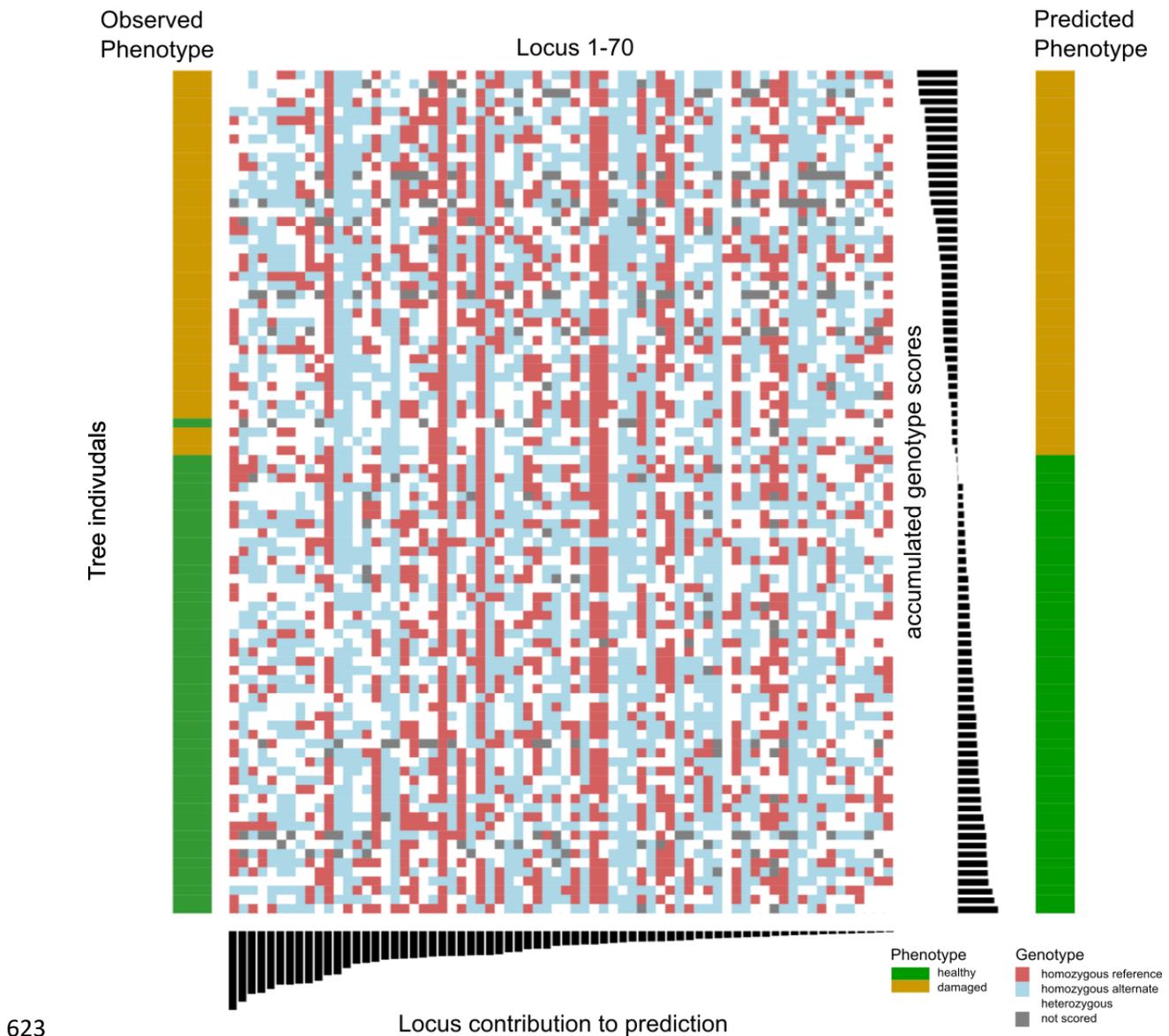


Figure 5. Combined results of SNP assay and discriminant analyses. The centre of the figure depicts the genotypogram of the SNP assay. Each column represents one of 70 loci, each row one of 92 beech individuals. The scored genotypes are colour-coded, with red squares = homozygous reference allele, light blue = homozygous alternate allele, white = heterozygous SNP, grey squares = locus could not be scored in the respective individual. The left bar indicates the observed phenotype for each tree individual with ochre rectangles for damaged, and green for healthy trees. Below the genotypogram, the relative contribution of each locus to the predictive model of the discriminant analysis is indicated, ordered from high to low. On the right side, first the genotype model scores for each individual are given, with the according predicted phenotype (ochre = damaged; green = healthy).

## Tables

Table 1. Genes with significantly associated SNPs. Given are the chromosome number (CHR), nucleotide position (position), the gene ID for *Fagus sylvatica* (gene), the UniProt ID of the closest match (UniProt ID), the name of the gene (name), the nucleotide base in the reference (ref DNA base), and the alternate base (alt DNA base), if applicable, the amino acid of the reference (ref AA) and the alternate base (non-synonymous change), functional change (effect) and the phenotype associated with the alternate base.

CHR	position	gene	UniProt ID	name	ref DNA base	alt DNA base	ref AA	non-synonymous change	effect	phenotype assoc. with alt base
1	40374762	1.g3851.t1	none		A	G	C	R	SH side chain > positive charge	healthy
10	32290645	10.g3914.t1	none		C	T	F	-	-	
11	20479628	11.g2467.t1	EXOS5_ORYSJ	Exosome complex exonuclease RRP46 homolog	T	C	T	A	polar > hydrophobic	damaged
11	23722307	11.g2832.t1	PCN_ARATH	WD repeat-containing protein PCN	A	G	I	V	hydrophobic > hydrophobic	damaged
12	13901034	12.g1695.t1	F4I5S1_ARATH	PB1 domain-containing protein tyrosine kinase	C	A	P	Q	hydrophobic > polar	damaged
	13901063				C	T	Q	stop	termination	damaged
	13901082				A	T	H	L	positive charge > hydrophobic	damaged
	13901094				T	A	I	N	hydrophobic > polar	damaged
2	43326571	2.g4736.t1	none		G	A	R	C	positive charge > SH side chain	damaged
3	31226940	3.g3590.t1	none		C	T	Q	stop	termination	healthy
4	34077017	4.g3980.t1	CKX1_ARATH	Cytokinin dehydrogenase 1	C	A	G	C	no side chain > SH side chain	damaged
5	16359587	5.g1807.t1	GDI2_ARATH	Guanosine nucleotide diphosphate dissociation inhibitor 2	T	C	P	-	-	
6	19865311	6.g2227.t1	NDUS7_ARATH	NADH dehydrogenase [ubiquinone] iron-sulfur protein 7, mitochondrial	T	C	D	-	-	
6	26383172	6.g2921.t1	none		T	C	A	-	-	
7	1493904	7.g177.t1	TLP10_ARATH	Tubby-like F-box protein 10	C	T	G	-	-	
7	20242023	7.g2350.t1	PRK4_ARATH	Pollen receptor-like kinase 4	A	G	P	-	-	
7	4504799	7.g1655			C	T	W	stop	termination	damaged

7	31456694	7.g3617.t1	LSH4_ARATH	Protein LIGHT-DEPENDENT SHORT HYPOCOTYLS 4	G	T	R	-	-		
7	33110000	7.g3816.t1	none		G	A	L	-	-		
7	4504813	7.g552.t1	none		G	A	G	-	-		
	4504831				C	T	L	-	-		
8	29295139	8.g3494.t1	VATC_ARATH	V-type proton ATPase subunit C	G	A	G	-	-		
9	25538827	9.g3080.t1	PPA14_ARATH	Probable inactive purple acid phosphatase 14	G	C	K	N		positive charge > polar	damaged
9	37955715	9.g4504.t1	TBL33_ARATH	Protein trichome birefringence-like 33	G	C	M	I		hydrophobic > hydrophobic	damaged

---