Genomic signals of local adaptation and hybridization in Asian white birch

Gabriele Nocchi[1,2#], Jing Wang[3#], Long Yang[4], Junyi Ding[5,6], Ying Gao[4], Richard J. A. Buggs[1,2*], Nian Wang[5,6,7*]

[1]School of Biological and Behavioural Sciences, Queen Mary University of London, Mile End Road, London, E1 4NS, UK

[2]Royal Botanic Gardens Kew, Richmond, TW9 3AB, Surrey, UK

[3]Key Laboratory for Bio-resources and Eco-environment, College of Life Science, Sichuan University, Chengdu, 610064, China

[4]Agricultural Big-Data Research Center and College of Plant Protection, Shandong Agricultural University, Tai'an 271018, China

[5]State Forestry and Grassland Administration Key Laboratory of Silviculture in Downstream Areas of the Yellow River, College of Forestry, Shandong Agricultural University, Tai'an 271018, China

[6]Mountain Tai Forest Ecosystem Research Station of State Forestry and Grassland Administration, College of Forestry, Shandong Agricultural University, Tai'an, 271018, China.

[7]State Key Laboratory of Crop Biology, Shandong Agricultural University, Tai'an, 271018, China

[#]Contributed equally to this work

[*]Author for correspondence:

Richard Buggs: r.buggs@kew.org

Nian Wang: nian.wang@sdau.edu.cn

Abstract

Disentangling the numerous processes that affect patterns of genome-wide diversity in widespread tree species has important implications for taxonomy, conservation, and forestry. Here, we investigate the population genomic structure of Asian white birch (*Betula platyphylla*) in China and seek to explain it in terms of hybridization, demography and adaptation. We generate whole genome sequence data from 83 individuals across the species range in China. Combining this with an existing dataset for 79 European and Russian white birches, we show a clear distinction between *B. pendula* and *B. platyphylla*, which have sometimes been lumped taxonomically. Genomic diversity of *B. platyphylla* in north-western China and Central Russia is affected greatly by hybridization with *B. pendula*. Excluding these hybridized populations, *B. platyphylla* in China has a linear distribution from north-eastern to south-western China, along the edge of the inland mountainous region. Within this distribution, three genetic clusters are found, which we model as long diverged with subsequent episodes of gene flow. Patterns of co-variation between allele frequencies and environmental variables in *B. platyphylla* suggest the role of natural selection in the distribution of diversity at 7,609 SNPs of which 3,767 were significantly differentiated among the genetic clusters. The putative adaptive SNPs are distributed throughout the genome and span 1,633 genic regions. Of these genic regions, 87 were previously identified as candidates for selective sweeps in Eurasian *B. pendula*. We use the 7,609 environmentally associated SNPs to estimate the risk of non-adaptedness for each sequenced *B. platyphylla* individual under a scenario of future climate change, highlighting areas where populations may be under future threat from rising temperatures.

Keywords

Asian white birch, landscape genomics, climate change, local adaptation, hybridization

Introduction

Analysis of the population genomic diversity of species can help us to understand their past and predict their future. Past processes affecting present allele frequencies include hybridization, range contraction and expansion, population size changes, gene flow among populations, and natural selection. Understanding the effects of natural selection is particularly important for future predictions as the extent to which alleles are locally adaptive has important implications for estimating population fitness under climate change (Sork et al., 2013; Savolainen et al., 2007; Aitken et al. 2008). Covariance between allele frequencies and environmental variables may be signatures of local adaptation but only if the effects of other past processes are accounted for (Sork et al., 2013; Wang & Bradburd, 2014). Hence, we need to gain the fullest possible understanding of all past processes affecting allele frequencies in populations before we can predict patterns of diversity that may be beneficial in future climates.

Recently, the genetic basis of local adaptation has been investigated for certain non-model species using high-throughput sequencing, high-resolution environmental data, and novel analytical methods that seek to account for the effects of neutral processes (Eckert et al., 2010; Ellegren, 2014). Environmental association analysis (EAA) correlates abiotic data with genomic data, such as SNP allele frequencies (Manel and Holderegger 2013; Sork et al. 2013). Such studies pinpoint SNPs showing signals of elevated differentiation among environments and showing significant correlations with one or more environmental variables (Rellstab et al., 2015), which are putative candidates involved in local adaptation (Hoban et al., 2016). Using EAA, we can also predict the possible degree of maladaptation of local populations to future climate change (Borrell et al., 2019; Rellstab et al., 2021; Rellstab et al., 2016).

The future of forest trees under climate change is critical for carbon sequestration, ecosystem conservation and timber production. Forest trees with wide geographic distributions over spatially heterogeneous environments often show local adaptation with adaptive traits polygenically

75 determined (Savolainen et al., 2007; Alberto et al., 2013). Due to their long generation times, forest
76 trees may show adaptational lag to rapid climate change (Browne et al., 2019; Keenan, 2015). The
77 genus *Betula* (birch) is a case in point, containing several tree species widespread in the Northern
78 Hemisphere, comprising key components of temperate forests (Ashburner and McAllister, 2013;
79 Wang et al., 2016; Wang et al., 2021). Understanding the genomic basis of adaptation in this genus
80 can help to predict their potential to survive under rapid climate change, helping to manage species
81 conservation (Borrell et al., 2019).

82 In this study, we selected Asian white birch (*Betula platyphylla*), which belongs to the subgenus
83 *Betula*, to understand its past demographical histories, current local adaptation, and future potential
84 to respond to climate change. Asian white birch is of great ecological and commercial value. It
85 commonly occurs in eastern Asia including the Russian Far East, China, and Japan. In China, it is
86 found in northwestern (NW), central (CE), northeastern (NE) and southwestern (SW) regions ranging
87 over 26 latitudinal degrees (Chen & Lou, 2019; Ashburner and McAllister, 2013). It is a pioneer
88 species, colonizing open habitats in grasslands, on roadsides and on the verge of forest (Ashburner
89 and McAllister, 2013). In the south it can exist at altitudes of over 3,700 meters whereas in NW China
90 it can grow by riversides in areas with much higher annual evaporation than precipitation.

91 Some taxonomists place *B. platyphylla* into the same species as silver birch, *B. pendula* (Ashburner &
92 McAllister, 2013), which is widespread in Europe and Russia. Others see them as sister species (Chen
93 et al., 2021; Wang et al., 2021; Salojärvi et al., 2017; Tsuda et al., 2017), with estimates of divergence
94 times differing between 36,000 years (Tsuda et al., 2017) and 2.6 million years (Chen et al., 2021).
95 Well assembled and annotated reference genomes now exist for both *B. pendula* (Salojärvi et al.,
96 2017) and *B. platyphylla* (Chen et al., 2021) but the latter was not available when we commenced the
97 present study. A large population genomic dataset is publicly available for *B. pendula*, analysis of
98 which identified two genetic clusters with some inter-mixed populations and showed that variants of
99 genes involved in photoperiodic responses were highly associated with local environmental

100 adaptation (Salojärvi et al., 2017). The genetic structure of *B. platyphylla* has been previously
101 investigated using ten microsatellite markers, showing five genetic clusters within its range, with
102 substantial gene flow among them (Chen & Lou, 2019).

103 In this present study, we generate whole genome sequence data for *B. platyphylla* across its range in
104 China to investigate its population genetic structure. We ask if *B. platyphylla* is differentiated from *B.*
105 *pendula*, and if so, the extent to which hybridization among them affects *B. platyphylla* populations.
106 We investigate the genetic structure within nonhybridized *B. platyphylla*, modelling how this has
107 been affected by population divergence, population size change, and gene flow. We also model the
108 environmental niche of *B. platyphylla* and how this may affect its range under future climates. We
109 use EAA to identify SNP markers in *B. platyphylla* associated with environmental clines after
110 controlling for population structure. We calculate the risk of non-adaptedness (RONA) (Rellstab et al.,
111 2016) for each individual under a scenario of climate of 2080-2100. Together, this study generates
112 new hypotheses regarding local adaptation in Asian white birch, which could be further tested in
113 common garden trials or trans-genetic analyses in future studies.

114

115 **Materials and Methods**

116 **Sampling**

117 Samples of Asian white birch (*B. platyphylla*) were collected from 74 naturally occurring populations
118 throughout its distribution in China between 2016 and 2019. Sample identifiers and GPS locations
119 are presented in Table S1. Dried cambial tissues were used for DNA extraction following a modified
120 CTAB protocol (Wang et al., 2013). One or two samples per population were selected for
121 resequencing, resulting in a total of 83 samples. Extracted DNA was assessed with 1% agarose gel
122 and sent to BerryGenomics (Beijing, China) for library preparation and whole genome re-sequencing.
123 Sequencing was performed on the Illumina NovaSeq6000 platform with 150 bp paired end read

124 sequencing with approximate 11Gbp data obtained for each sample (average depth of ~25x). We
125 also downloaded genomic read data from Salojarvi et al. (2017) for 78 *B. pendula* individuals from
126 most of this species' European and Asian geographic range and one *B. platyphylla* individual from
127 Russia.

128

129 **Trimming and SNP filtering**

130 The raw data was trimmed using Trimmomatic (Bolger et al., 2014) in paired-end mode with a
131 required quality of 30. Reads shorter than 90 bp were discarded. Clean reads of the sequenced
132 samples were aligned to a genome assembly of *B. pendula* (Salojärvi et al., 2017) using BWA-MEM
133 v.0.7.17-r1188 algorithm in BWA (v0.7.17) with default parameters (Li & Durbin, 2009). Reads with
134 non-specific matches were discarded. Alignments were converted from sequence alignment map
135 (SAM) format to sorted, indexed binary alignment map (BAM) files (SAMtools v1.8) (Li et al., 2009).
136 The MarkDuplicates tool from the Genome Analysis Tool Kit (GATK) (v 4.1.4.) was used to mark
137 duplicates (DePristo et al., 2011; McKenna et al., 2010). We filtered for biallelic SNPs genotyped in all
138 individuals with minor allele frequency greater than 0.05, quality by depth greater than 2, fisher
139 strand test less than 60, root mean square mapping quality greater than 40, mapping quality rank
140 sum test greater than -12, read position rank sum test greater than -8 and strand odds ratio less
141 than 3. We pruned the filtered SNP set by linkage disequilibrium ($r^2 > 0.4$) using the "indep-
142 pairphase" function of Plink (Chang et al., 2015) in windows of 50 and step of 5. We term this set of
143 SNPs the China dataset.

144 Using the same method to the above we also assembled a second SNP dataset that included
145 genomic read data from Salojarvi et al. (2017) for 78 *B. pendula* individuals from most of this species'
146 European and Asian geographic range, and one *B. platyphylla* individual deliberately included from
147 Russia. We term this set of SNPs the Eurasian dataset.

148

149 **Population structure**

150 We used *fastSTRUCTURE* (Raj, Stephens, & Pritchard, 2014) on the China dataset and also on the
151 Eurasian dataset to assign individuals to populations. We used a simple prior, 5-fold cross-validation
152 and number of ancestral populations (K) from one to ten. We used the python function *chooseK*,
153 which is the recommended tool for model selection in *fastSTRUCTURE* (Raj, Stephens & Pritchard,
154 2014), and the cross-validation profile, which identifies the model complexity for which the
155 prediction error is minimized, to assess the models generated. The function *chooseK* suggests two
156 values for K: one that maximizes the log-marginal likelihood lower bound (LLBO) of the dataset (K*?)
157 and captures strong structure, and another one (Køc) which reports the model components that
158 have a cumulative ancestry contribution of at least 99% and it is aimed at finding additional weak
159 underlying structure (Raj, Stephens, & Pritchard, 2014). Individual trees were assigned to
160 populations according to the value of the admixture coefficient (q) computed by *fastSTRUCTURE* at
161 the chosen model complexity: $q \geq 0.9$ to belong to a population, while individuals with $q < 0.9$ were
162 classified as admixed.

163 On the basis of the *fastSTRUCTURE* analysis of the Eurasian dataset, we excluded from the China
164 dataset any individuals that had over 10% assignment to *B. pendula* and ran *fastSTRUCTURE* only on
165 Chinese *B. platyphylla* based on 1,387,994 SNPs (after filtering for missing genotypes). Furthermore,
166 we used the program sparse non-negative matrix factorization (*snmf*), part of LEA package (Caye et
167 al., 2019), to perform a population differentiation test and identify F_{st} outlier SNPs between the
168 clusters identified for *B. platyphylla* in China. *snmf* computes fixation also in the presence of
169 admixed individuals in the sample (Martins et al., 2016) and performs a test similar to other F_{st}-
170 outlier approaches (Lotterhos & Whitlock, 2014). Multiple testing was controlled by converting the
171 p-values computed by *snmf* into q-values with the R package "qvalue" (Storey et al., 2015). Q-values

172 represent the expected false discovery rate (FDR) associated with a given p-value (Storey and
173 Tibshirani 2003). Finally, we selected candidate outlier SNPs with FDR < 1%.

174 Principal component analysis (PCA) was performed on the Eurasian dataset and the Chinese *B.*
175 *platyphylla* dataset using *Plink v2.0* (Chang et al., 2015).

176

177 **Demographic histories**

178 We inferred the demographic histories of the Chinese *B. platyphylla* dataset (excluding NW China)
179 for populations specified with K=3 and K=2 in the STRUCTURE analyses above (the relative merits of
180 these divisions reported in the Results and Discussion sections below) using a coalescent simulation-
181 based method implemented in fastsimcoal2 v 2.6.0.3 (Excoffier et al. 2013). The pairwise two-
182 dimensional joint SFSs (2D-SFSs) among the populations were constructed from the posterior
183 probabilities of sample allele frequencies using ANGSD v 0.934 (Korneliussen et al., 2014). A total of
184 28 models were evaluated for K=3 assuming divergence with or without gene flow and a constant or
185 changed population size after divergence (Figure S1; Table S2). Fourteen models specified that NE
186 diverged first, followed by the split of the CE and SW populations and 14 models specified the CE
187 population as of hybrid origin between the NE and SW populations (Figure S1). In addition, we ran
188 nine models for K=2 (Figure S2). The global maximum likelihood estimates for all demographic
189 parameters under each model were obtained from 60 independent runs, with 100,000 coalescent
190 simulations per likelihood estimates and 40 cycles of the likelihood maximization algorithm. The
191 models were compared based on the maximum value of likelihood over the 60 independent runs
192 using the Akaike's weight. The model with the maximum Akaike's weight value was chosen as the
193 optimal one. Confidence intervals were generated by performing parametric bootstrapping with 100
194 bootstrap replicates, and with 30 independent runs in each bootstrap. We used a mutation rate of
195 9.5×10^{-9} per site per year (Salojärvi et al. 2017) and a generation time of 15 years to convert the

196 coalescent scaled time into a speculative time scale in years. The best-fitting model was validated via
197 comparing the observed SFS to the predicted SFS based on the model (Figure S3).

198 In addition, we inferred the historical change in effective population sizes using Multiple Sequentially
199 Markovian Coalescent approach (MSMC v2.0.0) (Schiffels & Durbin, 2014) at K=3 for each of the
200 three populations, respectively. For each population, we randomly chosen seven individuals and
201 applied simple trio-phasing on any four individuals (eight haplotypes) using scripts provided in the
202 msmc-tools github repository (https://github.com/stschiff/msmc-tools). This resulted in 35 different
203 configurations for the MSMC analysis. We ran MSMC on each of the 35 configurations and estimated
204 medians and standard deviations of effective population sizes changes across time.

205

206 **Introgression test**

207 We used the ABBA-BABA tests to detect signals of hybridization (Green et al. 2010; Durand et al.
208 2011). We based these tests on all individuals from the SW, NW and CE populations, ten individuals
209 from the NE population and ten *B. pendula* individuals from Salojarvi et al. (2017). F_1 hybrids
210 detected by fastSTRUCTURE with a Q-score of approximately 0.5 were not included. We detected
211 asymmetries in the frequency of shared and derived variants (Green et al. 2010; Durand et al. 2011)
212 using different combinations of three populations, using *Betula occidentalis* from Salojärvi et al.
213 (2017) as an outgroup for all tests. The D and Z statistic were calculated for each combination using
214 function "Dtrios" in Dsuite 0.4-r38 (Malinsky et al., 2021). A block jack-knifing was performed to
215 obtain an associated p value for the test of whether D statistic deviates significantly from zero. An
216 average Z score larger than 3.09 (equal to a p value < 0.001 of D statistic) was regarded as a
217 significant signal of introgression.

218

219 **Population diversity and differentiation**

220 We estimated linkage disequilibrium decay along the *B. platyphylla* genome with the tool
221 *PopLDdecay* (Zhang et al., 2018) separately for the different populations identified for *B. platyphylla*
222 in China by *fastSTRUCTURE* at K = 3 (Figure 3). For this calculation we used the SNP set prior to
223 linkage disequilibrium pruning but filtered to retain SNPs with minor allele frequency > 0.05. After
224 excluding individuals admixed between the three populations (admixture coefficient < 0.9 in
225 *fastSTRUCTURE* at K = 3), Weir and Cockerham F_{st} (Weir & Cockerham, 1984) was calculated
226 between the three *B. platyphylla* populations on a per SNP site basis with the *vcftools* function *weir-*
227 *fst-pop* (Danecek et al., 2011). We calculated nucleotide diversity π (Nei & Li, 1979) within each
228 population using the *vcftools* (Danecek et al., 2011) function *window-pi* in non-overlapping windows
229 of 5 kb. For the π calculation we used the SNPs set prior to minor allele frequency filtering and
230 linkage disequilibrium pruning (28,058,885 SNPs in total).

231

232 **Environmental niche modelling**

233 We used environmental niche modelling (ENM) to characterize the present habitat and predict the
234 future distribution of non-hybrid *B. platyphylla* in China and to identify the climatic variables
235 influencing this species distribution. ENM analysis was performed by using the samples' geographic
236 locations in this study and observation records of this species from 1970 to 2010 in China, sourced
237 from the Global Biodiversity Information Facility (http://www.gbif.org). To mitigate the effects of
238 spatial autocorrelation, we removed records within 0.2 degrees (approximately 22-25 km) of one
239 another, which resulted in a total of 138 presence locations: 66 records were *B. platyphylla* from the
240 83 individuals sampled (excluding admixed *B. pendula* - *B. platyphylla* individuals) and another 72
241 records were from GBIF (excluding NW China as all the trees that we had sampled there were
242 hybrids – see Results). We downloaded 19 bioclimatic variables related to temperature and
243 precipitation from the WorldClim database (www.worldclim.org) at 1km resolution (Hijmanset al.,
244 2005) for the period 1970-2000 representing "current climate", as this is the estimated time period

245 of establishment of the sampled trees and also corresponds to the time that most GBIF observations
246 included in the model were recorded. We downloaded elevation data at the same resolution and
247 used it to compute slope, aspect and four additional terrain related characteristics (topographic
248 position index, terrain ruggedness index, terrain roughness and water flow direction) using the
249 *raster* R package (Hijmans & Etten, 2012). In order to avoid overfitting, we selected environmental
250 variables with Pearson's correlation coefficient < 0.7, preferring annual rather than monthly or
251 quarterly values, which resulted in 11 variables retained for the ENM (Table S4, Figure S4-5). We
252 assembled eight further datasets with the same 11 variables under four Shared Socioeconomic
253 Pathways (SSPs) defined by the Intergovernmental Panel on Climate Change sixth Assessment
254 (Masson-Delmotte et al., 2021) at each of two future time points (2041–2060 and 2081–2100). The
255 ENMs were generated using Maxent (Phillips et al., 2006) with 50 subsampled replicate 5000
256 iterations runs with 20% of observations left-out for cross-validation. Variables' importance was
257 assessed with jack-knife tests and multiple models were generated and evaluated using a test of
258 omission rate and area under the receiver operating characteristic curve (AUC). An environment
259 suitability threshold was defined by "maximum training sensitivity plus specificity," which optimizes
260 the trade-off between commission and omission errors (Liu et al., 2016).

261

262 **Genome environment association analysis: identification of adaptive SNPs**

263 We used *LFMM2* (Caye et al., 2019) to test for associations between environmental variables and
264 SNP allele frequencies in 71 Chinese non-hybrid *B. platyphylla* individuals. *LFMM2* performs
265 multivariate linear regressions to evaluate the association between a response matrix,
266 corresponding to SNP frequencies of individuals, and a matrix of environmental variables. It
267 combines the environmental fixed effects with latent effects, which are unobserved confounding
268 effects due to population structure (Frichot et al., 2013; Caye et al., 2019). *LFMM2* has a two steps
269 approach: first it estimates the latent factors and then tests for the genotype-environment

270 associations. Each SNP locus is tested separately using the latent scores as covariates to control for
271 demographic history and population structure. The null hypothesis is that the environmental
272 variables have no effects on the SNP frequency, which is tested using a student distribution with
273 n−K−1 degrees of freedom (Caye et al., 2019). *LFMM2* requires to specify the number (K) of latent
274 factors to include in the model which was chosen on the basis of *fastSTRUCTURE* cross-validation
275 schemes and a Tracy-Widom test on the PCA eigenvalues. We chose the more conservative K = 3
276 (see Results section) to run *LFMM2*, including all the 11 uncorrelated environmental variables used
277 in the ENM (Table S4). Multiple testing was controlled by converting the p-values computed by
278 *LFMM2* into q-values with the R package "qvalue" (Storey et al., 2015). Finally, we selected
279 candidate adaptive SNPs with FDR < 1%.

280

281 **Characterization of candidate adaptive SNPs**

282 We performed two further principal component analyses of the Chinese *B. platyphylla* samples with
283 *Plink* (Chang et al., 2015): one including only the putative adaptive SNPs (detected by *LFMM2* (K = 3)
284 with FDR < 1%) and another only including a similar number of putatively neutral SNPs. The neutral
285 set of 7,500 SNPs was generated by selecting SNP loci at random from those that reported a q-value
286 above 0.4 across all environmental variables in the *LFMM2* analysis. Weir and Cockerham F_{st}
287 between the three *B. platyphylla* populations was re-computed based on the adaptive and neutral
288 SNP sets with the *vcftools* function *weir-fst-pop* (Danecek et al., 2011) after excluding admixed
289 individuals (admixture coefficient < 0.9 in *fastSTRUCTURE* at K = 3).

290 In order to functionally annotate our putative adaptive SNPs, we performed a gene ontology (GO)
291 annotation of the whole *B. pendula* reference gene set provided in Salojärvi et al. (2017), retrieving
292 GO terms associated with the coding regions (Ashburner et al., 2000) using *Omics Box* (Conesa &
293 Götz, 2008). We used *Omics Box* default functional annotation workflow which incorporates BLAST,
294 run on the Viridiplantae non-redundant database, and Interproscan (Zdobnov & Apweiler, 2001). We

295 then identified the genic regions overlapping our candidate adaptive SNPs with *bedtools* v2.28.0
296 (Quinlan & Hall, 2010) and performed a functional enrichment analysis of this adaptive gene set
297 against the fully annotated reference gene set using Fisher's exact test in *Omics Box* (Conesa & Götz,
298 2008). We applied multiple testing corrections (FDR < 5%) and reduced the resulting significantly
299 enriched GOs to the most specific term in the hierarchy for all three ontology levels: biological
300 process, molecular function, and cellular component (Ashburner et al., 2000).

301

302 **Adaption to future climate**

303 We carried out risk of non-adaptedness (RONA) analysis using a slight modification of the method of
304 Rellstab et al. (2016), implemented in the software *pyRona* (Pina-Martins et al., 2018). RONA
305 represents an estimate of the average change in allele frequencies at climate associated SNP loci
306 required in a population to cope with future climatic conditions and is calculated separately for each
307 environmental variable (Rellstab et al., 2016). RONA uses simple linear regressions of the selected
308 current climate environmental variable and the alternative allele frequencies of the candidate loci
309 identified in genome-environment association analyses (Rellstab et al., 2016). The regression
310 coefficients of the environmentally associated loci are then used to predict the allele frequencies of
311 the adaptive loci at a future climate value for the chosen environmental variable (Rellstab et al.,
312 2016). The main improvement in the implementation by Pina-Martins et al. (2018) is that the RONA
313 for an environmental variable is given by the weighted mean RONA of all relevant SNP loci for that
314 environmental variable, weighted by the r^2 value of each locus regression, while the original method
315 (Rellstab et al., 2016) uses unweighted means. We computed RONA on an individual sample basis for
316 the seven uncorrelated environmental variables that are expected to change in the future, therefore
317 excluding elevation and its derived measures. We used as future climate the profile ssp370 for 2080-
318 2100 (Figure 4C). For each environmental variable, RONA was calculated by using the candidate
319 adaptive SNP loci identified with *LFMM2* (FDR < 1%). For each individual tree we then calculated

320 weighted mean RONA, giving each environmental variable RONA a weight equivalent to the
321 contribution reported by that variable in the ENM. Furthermore, we reported the maximum RONA
322 for each individual, out of the seven variables for which it was calculated.

323

324 **Results**

325 **SNP discovery**

326 After filtering, the number of reads retained for each sample in the China dataset that we generated
327 ranged between 66,016,635 and 134,129,672 and between 91.5% and 96.7% of these reads per
328 individual were mapped to the reference genome (Table S1). SNPs filtering of the 83 individuals
329 resulted in 1,497,547 unlinked (r^2< 0.4) SNPs. After filtering, the number of reads retained for each
330 sample in the Salojarvi et al. (2017) dataset ranged between 16,506,088 and 237,845,424 and
331 between 82.9% and 97.2% of these reads per individual were mapped to the reference genome. The
332 combined Eurasian dataset resulted in 278,717 SNPs after filtering for missing genotypes. When the
333 China dataset was restricted to pure *B. platyphylla* (71 individuals in SW, CE, and NE China), we were
334 left with 1,387,994 SNPs.

335

336 **Population structure, diversity, and differentiation**

337 For the China dataset *fastSTRUCTURE* analysis with K = 2 separated the samples in the NW
338 population (Xinjiang) from the other populations. At K = 3, the SW and NE populations were further
339 separated and at K = 4 the CE population was revealed (Figure 1). The model that maximized the log-
340 marginal likelihood lower bound (LLBO) of the data and that best explained additional weak
341 underlying structure was K = 3, as suggested by the function "*chooseK*" (Figure S6).

342 For the Eurasian dataset at K = 2, one cluster contained the *B. pendula* individuals from Salojarvi et
343 al. (2017) and the other cluster contained 71 of the 83 individuals that we sequenced from China

344 and the Russian *B. platyphylla* individual from Salojarvi et al. (2019) (Figure 2). The 12 individuals we
345 sampled in the NW population were shown as hybrids between *B. pendula* and *B. platyphylla* (Figure
346 2A). At K = 3, the pattern was similar but a SW *B. platyphylla* clustered separately from a NE *B.*
347 *platyphylla* cluster (Figure 2B). At K = 4, the only difference was that Irish and two Finnish *B. pendula*
348 samples showed an unknown component (Figure 2C). The *fastSTRUCTURE* model that maximized the
349 log-marginal likelihood lower bound (LLBO) of the data and that best explained additional weak
350 underlying structure was K = 4, as suggested by the function choose K (Figure S7). The LLBO curve
351 and the cross-validation profile showed that the marginal likelihood plateaued at K = 2, and so did
352 the prediction error, which stayed within 1 standard error from K = 2 with added complexity (Figure
353 S7A-B). These separations suggested by *fastSTRUCTURE* were reflected in the PCA (Figure S8).

354 In *fastSTRUCTURE* restricted to the 71 non-hybrid *B. platyphylla* individuals in China, the model that
355 maximized the log-marginal likelihood lower bound (LLBO) of the data was K = 2 and that best
356 explains additional weak underlying structure was K = 3, as suggested by the function "*chooseK*"
357 (Figure S9). The *fastSTRUCTURE* cross-validation score profile showed that the prediction error is
358 lowest at K = 3 (Figure S9B). At K = 2 the NE population was separated from the SW population and
359 the individuals in the CE population appeared admixed between these two populations, while at K =
360 3 the CE population was distinct (Figure 3A and 3B).

361 In PCA of the 71 pure *B. platyphylla* individuals, PC1 explained ~33% of the total variance and
362 separated the NW, CE, and SW populations (Figure S10). PC2 (15% of total variance) also separated
363 the three populations while PC3 (8% of total variance) scattered the individuals of the NE
364 population. The Tracy-Widom test identified three significant principal components ($p < .05$) and the
365 explained variance levelled-off after PC3 (Figure S10D).

366 Due to the differentiation between the individuals located in central China and those in the North-
367 East shown by the PCA (Figure S10) and the recommendation by *fastSTRUCTURE* cross-validation
368 scheme (Figure S9B), we assigned individuals to populations according to the admixture coefficients

369 computed with *fastSTRUCTURE* at K = 3. In total, ten individuals were assigned to the central
370 population (CE), seven to the south-western population (SW), 46 to the north-eastern population
371 (NE) and the remaining eight individuals were classified as admixed (Figure 3B-C). Based on this
372 assignment, *snmf* identified 17,218 outlier SNPs with FDR < 1% (Figure S11).

373 Linkage disequilibrium decays rapidly in *B. platyphylla* populations and reaches background levels at
374 approximately 50 kb (Figure S12). Assessment of pairwise F_{st} between populations showed that the
375 highest levels of differentiation were between the NE and SW populations (mean F_{st} = 0.10) or the
376 CE population and SW populations (mean F_{st} = 0.11) (Figure S13). On the other hand, the CE and NE
377 populations reported significantly lower mean pairwise F_{st} (0.05) (Figure S13).

378 Nucleotide diversity π in windows of 5kb showed a slight increase in populations from south to north
379 geographically, reporting a mean of 0.0057 (~0.6%) in the SW population, 0.0075 (~0.7%) in the CE
380 population and 0.0083 (~0.8%) in the NE population (Figure S14-15).

381 **Species population size and separation history**

382 Assuming that there are three populations of *B. platyphylla* in NE, CE and SW China (K=3), and the
383 NE population was the first to diverge, the best-supported fastsimcoal2 model (Figure 3D; Table S2)
384 suggested that the NE population diverged from the common ancestral population of CE and SW at
385 2.804 Mya (95% CI: 2.892–3.506 Mya) (Figure 3D). The NE population initially had a small population
386 size. The CE and SW populations diverged 2.797 Mya (95% CI: 2.294–2.768 Mya; Figure 3D). The SW
387 population underwent a slight bottleneck at 2.787 Mya (95% CI: 2.129–2.489 Mya) and experienced
388 a slight population expansion at 0.812 Mya (95% CI: 0.477–0.706 Mya) (Figure 3D). The estimated
389 present day effective population sizes (Ne) are 3.3×10^5 (95% CI: $3.0–3.4 \times 10^5$) for NE, 1.2×10^5 (95%
390 CI: $1.6–2.2 \times 10^5$) for CE and 0.7×10^5 (95% CI: $1.6–2.2 \times 10^5$) for SW. The best-supported model
391 detected low levels of ancient gene flow between the populations soon after their divergence, no
392 gene flow between 2.797 Mya and 0.812 Mya, and recent gene flow between the NE and CE
393 populations and between the SW and CE populations (Figure 3D). We also ran fastsimcoal2 models

394 assuming K=2 and the best model suggested an older initial divergence than we found with K=3 and
395 an early bottleneck in the SW population only. Patterns of gene flow and current Ne were similar to
396 those of the K=3 model.

397 (Figure S2, Table S3).

398 MSMC results suggested that the three populations experienced severe bottlenecks (2.51-2.89
399 million years ago) followed by population expansion during the period of 26.1-42.6 thousand years
400 ago (Figure 3E). The NE lineage experienced the least population reduction (Ne, 11235.88 ± 204.75)
401 whereas the SW lineage underwent the most severe population decline (Ne, 6152.70 ± 265.12)
402 (Figure 3E).

403

404 **Hybridization between *B. platyphyllya* and *B. pendula***

405 The ABBA-BABA tests showed a high degree of introgression from *B. pendula* into the NW *B.*
406 *platyphylla* population (mean D-statistic = 0.17, mean Z-score = 15.02), which was also identified as
407 having mixed ancestry in the fastSTRUCTURE analysis (Figure 2; Figure 3F). The ABBA-BABA results
408 also indicatpotential gene flow between the NE, CE and SW populations (Figure 3F).

409

410 **Environmental niche modelling**

411 Environmental niche models were built based on eleven uncorrelated environmental variables
412 (Figure S4-5; Table S4) and were based on our *B. platyphylla* samples' locations and GBIF
413 observations of this species (Figure 4).

414 Jack-knife tests suggested that the variables "aspect" and "water-flow direction" did not play a
415 relevant contribution either singly or in combination, so they were excluded (Figure S16-18). The
416 final Maxent model with the nine remaining variables reported high mean test AUC (0.913 ± 0.017)

417 and low mean test omission rate (0.13, $p < .001$) at a maximum training sensitivity plus specificity
418 logistic threshold of 0.2 (Figure 4A-B; Figure S19-20). Two variables, "annual precipitation" and
419 "mean temperature of wettest quarter", together contributed to > 50% of the predicting
420 performance of the model, defined as the increase in regularized gain added (or subtracted) to the
421 contribution of the corresponding variable, over each model iteration (Table S5).

422 Response curves for the final Maxent model are available in Supporting Information (Figures S28-
423 S36). Future projections show an overall contraction of the *B. platyphylla* habitat throughout China,
424 particularly noticeable in central China and in the north-east (Figure 4C). Currently suitable
425 environments located at higher elevations seem to contract less in the future (Figure 4C). The
426 predicted habitat reduces between 19 and 30 % in 2040-2060, and between 17 and 34 % in 2080-
427 2100 (Figure 4C), compared to the present. In the most adverse scenario (ssp370), *B. platyphylla*
428 habitat in China may be reduced by approximately a third of its current extent by 2080-2100 (Figure
429 4C) and the NE and SW populations will be spatially separated by a large uninhabitable region.

430

431 **Genome-environment association analysis: identification and characterization of putatively**
432 **adaptive SNPs**

433 We choose a q-value threshold of 0.01 (FDR < 1%) to select candidate associations in the LFMM2
434 analysis on 1,387,994 SNPs ($r^2 < 0.4$) restricted to the 71 non-hybrid *B. platyphylla* individuals
435 identified with *fastSTRUCTURE* (i.e. excluding the individuals in NW China).

436 In total, 7,609 SNPs showed significant associations with one or more environmental variables in
437 *LFMM2* (K = 3) using our criteria (Figure S21), for a total of 11,304 associations. In more detail: 4,643
438 SNP were associated to one environmental variable, 2,424 SNPs to two variables, 384 SNPs to three
439 variables, 133 SNPs to four variables, 21 SNPs to five variables and four SNPs to six variables. Of the
440 7,609 putative adaptive SNPs identified with *LFMM2*, 3,767 were shown to be more differentiated

441 among the NE, CE and SW populations than was expected due to drift in our *snmf* F_{st} outlier test
442 (FDR < 1%).

443 The number of the putative environmentally associated SNPs detected with LFMM2 differed among
444 the variables tested (Figure S22). "Isothermality" and "mean diurnal range" had 5,679 and 2,794
445 putative SNP-environment associations identified, respectively. These were followed by "annual
446 precipitation" (1,059 SNPs hits), "precipitation of driest quarter" (735 SNPs hits), "annual mean
447 temperature" (391 SNPs hits), "flow direction" (263 SNPs hits), "mean temperature of wettest
448 quarter" (227 SNPs hits), "TPI" (104 SNPs hits), "slope" (49 SNPs hits) and "precipitation seasonality"
449 (3 SNPs hits). The variable "aspect" did not have any SNP association with an FDR < 1% in *LFMM2*.
450 There was no significant correlation between the number of identified SNP association per variable
451 and the percentage of contribution of each variable to the ENM (Pearson's test $p > 0.05$).

452 The candidate adaptive SNPs were spread across the entire genome and present on all linkage
453 groups in the *B. pendula* reference genome (Figure S22). The chromosome with the largest number
454 of candidates is chromosome 2 (738 SNPs), while the lowest number was recorded for chromosome
455 6 (344 SNPs). The number of SNPs identified per chromosome does not appear correlated with
456 chromosome size (Pearson's test $p > 0.05$). A hotspot was visible in the second half of chromosome 7
457 (Figure S22).

458 The PCA based only on the 7,609 adaptive SNPs showed a steep discrepancy in the variance
459 explained by PC1 compared to that of the other PCs (Figure S23). Furthermore, PCA based on only
460 putatively adaptive SNPs did not show a clear separation between the central and north-eastern
461 population in the first three PCs, while it separated the south-western population (Figure S23).
462 When using only 7,500 putatively "neutral" SNPs in PCA, the separation between populations was
463 slightly weaker compared to that observed with the whole data set, and individuals in each
464 population were more scattered along the PCA axes (Figure S24). The pairwise F_{st} patterns of
465 differentiation between populations based only the 7,609 adaptive SNPs and only on neutral loci

466 reflected those observed with the whole SNPs set, however loci under selection showed
467 substantially higher F_{st} compared to those recorded for neutral markers and for the whole SNP set
468 (Figure S25-26).

469 We identified 1,633 mRNA regions in the *B. pendula* reference genome (Salojärvi et al., 2017)
470 containing at least one candidate adaptive SNP,. A functional enrichment analysis of these genic
471 regions reported significant hits (FDR < 5%) at all three gene ontology levels: biological process (3
472 hits), cellular component (1 hits) and molecular function (3 hits) (Figure S27).

473

474 **Risk of non-adaptedness to future conditions (RONA)**

475 We calculated RONA for seven environmental variables projected under the future climate profile
476 ssp370, identified as the worst-case scenario for *B. platyphylla* for the years 2080 – 2100 in terms of
477 habitat reduction (Figure 5; Table S6). The most represented environmental variables in terms of
478 number of associations in the RONA computation were "isothermality" (5,679 SNPs, mean r^2 =
479 0.459), "mean diurnal range" (2,794 SNPs, mean r^2 = 0.0157) and "annual precipitation" (1,059 SNPs,
480 mean r^2 = 0.0976) (Table S6). The variables that reported higher average r^2 of adaptive SNPs were
481 "isothermality" (5,679 SNPs, mean r^2 = 0.459), "annual mean temperature" (391 SNPs, mean r^2 =
482 0.195), and "mean temperature of wettest quarter" (227 SNPs, mean r^2 = 0.185) (Table S6).

483 There was large variability in the expected allele frequency changes required to match future
484 conditions across environmental variables. "Annual mean temperature" and "mean temperature of
485 wettest quarter" showed notably higher mean individual RONA compared to all the other variables,
486 being 0.1773 and 0.2314 respectively (Table S6).

487 Weighted mean RONA scores were overall relatively low across individuals, with mean of 0.08 (sd =
488 0.03) and maximum of 0.1455, reported for an individual in the north-eastern population (HTY15)
489 (Figure 6A; Table S7). The maximum RONA scores showed greater variability across individuals, with

490 mean of 0.25 (sd = 0.09) and maximum of 0.3956, recorded for an individual in the north-eastern
491 population (JY23, "mean temperature of wettest quarter") (Figure 6B; Table S7).

492 There were some differences between the three Chinese *B. platyphylla* populations in regard to the
493 environmental variables that will require greater allele frequency changes (in ssp370 at 2080 – 2100)
494 (Table S8). Although all populations showed similar RONA patterns across the environmental
495 variables tested, with the largest mean RONA (averaged over the individuals of each population)
496 reported for temperature derived variables ("annual mean temperature" and "mean temperature of
497 wettest quarter") rather than precipitation-derived variables (Table S8), the south-western
498 population, had much higher mean RONA for "precipitation seasonality" (0.21) compared to the
499 other two populations that had relatively low mean RONA for this specific environmental variable
500 (Table S8).

501

502 **Discussion**

503 **Pan Eurasian population genetic structure of white birch**

504 Our study clearly differentiated between *B. pendula* and *B. platyphylla* across Eurasia, showing
505 strong separation between the two in PCA (Figure S8) and *fastSTRUCTURE* (Figure 2) analyses. We
506 found evidence for extensive hybridization between these two species in NW China, and to a lesser
507 extent in central Russia. Populations in NW China that were initially identified as *B. platyphylla*
508 turned out to be mainly *B. pendula* with introgression from *B. platyphylla*. A previous microsatellite
509 study of *B. pendula* and *B. platyphylla* across Eurasia (Tsuda et al., 2017) showed admixture between
510 the two species in Siberia, but in Asia this study only included populations from Russia and Japan. In
511 a microsatellite study of white birches in China, Chen & Lou (2019) found populations in northwest
512 China to be highly differentiated from other populations, and to be of high genetic diversity.
513 However, as no *B. pendula* were included in their study, they could not pick up a signature of

514 hybridization in these populations. Instead, they attributed their high differentiation and diversity to
515 a northern glacial refugium for white birch in the Altay Mountains. If we had not co-analysed our
516 data with that of Salojärvi et al. (2017), we might also have concluded that the populations in NW
517 China were a divergent lineage, rather than introgressed. In this study, we removed individuals from
518 the NW population from our downstream analyses on *B. platyphylla*.

519

520 **Genetic structure of white birch within China and signals of local adaptation**

521 With the NW individuals excluded, *B. platyphylla* in China has a linear distribution from north-
522 eastern to south-western China, along the edge of the inland mountainous region. Occurrence data
523 and environmental niche models suggest that there is currently a single contiguous region currently
524 habitable by *B. platyphylla*. Within our *B. platyphylla* samples, we found three genetic clusters with
525 some admixed individuals between them. The three clusters seem to mirror similar phylogeographic
526 breaks in many other plant species (Zhout et al., 2021; Ye et al., 2017; Fan et al., 2016; Bai et al.,
527 2016) and our interpretation of *B. platyphylla* population structure is similar to that of Chen & Lou
528 (2019). Our fastsimcoal results suggest that the three *B. platyphylla* populations exchanged genetic
529 material soon after their divergence and recently, but endured a long period of isolation in between
530 these contacts (Figure 3D). The absolute date estimated from fastsimcoal suggests that this period of
531 isolation between the *B. platyphylla* populations was from 2.8 Mya to 0.8 Mya. This is hard to fit
532 with estimates of the time of speciation between *B. platyphylla* and *B. pendula* which different
533 studies place between 36,000 years (Tsuda et al., 2017) and 2.6 million years (Chen et al., 2021) ago.
534 We note, however, that our demographic simulation results are provisional as it is hard to
535 distinguish among the many factors that affect patterns genetic variation (Johri et al., 2022), and our
536 assumptions about mutation rate and generation time may be incorrect. We cannot therefore
537 exclude the possibility that our simulations are picking up signals of *B. platyphylla* population
538 isolation due to separate refugia in recent glaciations.

539 Consistent with a previous phylogeographic study of Asian white birch, showing that microsatellite
540 diversity of Asian white birch increased with latitude (Chen & Lou, 2019), our results confirmed that
541 genomic diversity in the NE is higher than that in CE which is higher than that in SW (Figure S14-15).
542 This favours the scenario of separate glacial refugia at different latitudes, rather than a Europe-like
543 postglacial expansion from southern refugia (Birks & Willis, 2010; Hewitt, 1999). This pattern may
544 also reflect less severe geographical barriers to pollen movement in northern China compared to the
545 south-western Qinghai-Tibet Plateau. It might have been expected that higher diversity in the NE
546 population is due to admixture from *B. pendula*, but we find no evidence for this.

547 Ecological niche models indicated that mean temperature of wettest quarter and annual mean
548 precipitation are the most important predictors of *B. platyphylla* distribution and both north-eastern
549 China and the south-western Tibetan plateau currently provide the most suitable environments for
550 this species in China. Future climate projections show an overall decline of the species range and
551 environmental suitability throughout of its current distribution in China, with suitable habitats
552 persisting mostly in areas of higher elevation (Figure 4) and the central China habitable area
553 vulnerable to being lost.

554 Our PCA based only on adaptive SNPs showed that the NE and CE populations formed an
555 overlapping cluster whereas in PCA based on neutral SNPs the two populations were separated
556 (Figure S23-24). This is possibly due to the relatively lower diversifying selection acting on the two
557 populations and the existence of gene flow which can counteract the impact of diversifying selection
558 (Guichoux et al., 2013; Nosil et al., 2009). F_{st} analyses suggested that the level of differentiation is
559 much higher at loci under selection rather than neutral loci, as expected, and the SW populations
560 appears to be highly differentiated from the others at adaptive loci (Figure S25).

561 Functional enrichment analysis of mRNA intersecting the identified adaptive SNPs suggested that
562 these genes were mostly enriched for growth and environmental stress response (Figure S27),
563 although further studies and functional experiments are needed to confirm these findings. Among

564 the significantly enriched categories, the most interesting in view of climate adaptation were:
565 “regulation of response to stimulus”, “DNA helicase activity” which are enzymes involved in DNA
566 repair and have a crucial role as caretakers of the plant genome against environmental damages
567 (Sami et al., 2021) and “inorganic cation transmembrane transporter activity”, as potassium, the
568 most abundant inorganic cation in plant cells, it is essential for plant growth and development and it
569 has been shown to have a major role in resistance to drought, salinity and fungal infections (Sharma
570 et al., 2013). The 1,633 identified adaptive genes also included several light-response and growth-
571 related genes, according to our OmicsBox annotation, as well as members of all the three GO
572 categories significantly enriched in the putative genes under selection detected by Solojarvi et al.
573 (2017): transmembrane receptor protein tyrosine kinase, peptidyl-histidine phosphorylation, and
574 axis specification. Furthermore, 87 of the 1,633 genic regions detected in this study were identified
575 as putative selective sweeps in *B. pendula* (Salojarvi et al., 2017). Even though it would be interesting
576 to explore all the identified genes singularly, this goes beyond the extent of this work and further
577 experimental validation would still be necessary to confirm their role in growth and environmental
578 stress response.

579 In line with similar studies in other species (Jordan et al., 2017; Rellstab et al., 2016; Borrell et al.,
580 2019; Pina-Martins et al., 2018), local adaptation appears to be a highly polygenic in *B. platyphylla*,
581 therefore it is likely that maintaining the standing variation and adaptive diversity may be a better
582 solution to aid future climate adaptation, rather than focusing on raising the frequencies of a specific
583 set of adaptive SNPs, as it was proposed previously (Jordan et al. 2017). Two variables, “mean
584 diurnal range” and “isothermality” reported a substantially larger number of significant SNPs
585 association compared to the other climatic variables (Figure S22), which may have arisen due to the
586 spatial correlation of these environmental variables and genotypes across the sampling range
587 therefore these results should be interpreted with caution and further validation is necessary,
588 however we cannot exclude and highly polygenic adaptation to these variables with many SNPs with
589 low effects involved. We did not identify any significant correlation between the number of SNPs

590 associated with each climate variable, and their percentage contribution in our ENM, unlike some
591 other studies (Borrell et al., 2019; Rellstab et al., 2016). However, we note that it is not a logical
592 necessity as the variables with higher discriminatory power in the ENM could be limiting species
593 ranges either because they lack adaptation (Borrell et al., 2019), or contrarily they could be
594 subjected to strong adaptation but at a global level throughout the species range, rather than locally.
595 Therefore, the lack of variation in these variables across the sampling range prevents the detection
596 of these adaptive signals with methods designed for local selection, such as *LFMM2*.

597 To investigate how the samples of *B. platyphylla* will respond to rapid future climate change, we
598 calculated RONA for the 71 *B. platyphylla* individuals according to the implementation by Pina-
599 Martins et al. (2018). Overall, RONA was relatively low across the environmental variables tested
600 (Figure 5), with the exception of "annual mean temperature" and "mean temperature of wettest
601 quarter", very likely reflecting the larger relative projected change of temperature derived variables
602 compared with that of environmental variables related to precipitation throughout China. The
603 expected allelic frequencies changes are similar to those reported in other studies (Jordan et al.,
604 2017; Pina-Martins et al., 2019; Rellstab et al., 2016). Weighted mean RONA incorporated the ENM
605 with EAA and generally remains low and it is always below 0.2 (Figure 6A). We also reported the
606 maximum RONA per individual out of the seven variables tested (Figure 6B) because even if a
607 population, in this case an individual, has "low" average RONA for a given future projection, it might
608 still be its highest RONA that will determine how much it will need to shift its allelic frequencies in
609 order respond to future selective pressure (Pina-Martins et al., 2018). Maximum RONA across
610 individuals showed significantly higher values, as expected. Interestingly, the maximum RONA plot
611 also showed particularly high values for four individuals in the south-western population located at
612 the very southern edge of this population distribution, reported for precipitation seasonality which is
613 expected to have a much larger relative change at these locations (Figure 6B). Overall, both

614 weighted mean and maximum RONA match well with our ENM for ssp370 2080-2100, with lower
615 RONA values reported within the inferred projected suitable environment (Figure 6).

616 The RONA method is of course a huge simplification, and its limitations are accurately described in
617 the original publication (Rellstab et al., 2016), therefore our results should be viewed with caution. It
618 should also be noted that RONA was originally designed to be used with population allelic
619 frequencies rather than individuals, but given the nature of our sampling, using individuals was to
620 our knowledge the best approach. Grouping the individuals in populations for the RONA calculation
621 would have been problematic as it would have required averaging climatic conditions across
622 different locations. This would have led to inaccuracy due to the broad geographic range of our
623 study and the high heterogeneity of some of the environmental variables tested.

624 It is not straightforward to determine when RONA is high enough to result in a lag between allele
625 frequencies and adaptation. A previous study on *Fagus sylvatica* have observed allele frequencies
626 changes of 0.1 – 0.2 per decade (Jump et al., 2006), but this estimate was based on amplified
627 fragment length polymorphism (AFLP) molecular markers rather than SNPs. A more recent study on
628 Swiss stone pine, *Pinus cembra*, based on SNPs markers, investigated allele frequencies shifts over
629 time by looking at the differences in SNPs frequencies between adults and juvenile cohorts at seven
630 locations in Switzerland, and validated the observed shifts with forward-in-time simulations
631 (Dauphin et al., 2020). This analysis reported an average rate of allele frequency shifts of 1.26×10^{-2}
632 per generation (*P. cembra* generation time = 40 years) for neutral SNPs, and a slightly lower estimate
633 for adaptive SNPs. Based on this estimate we can hypothesize that SNPs frequencies changes < 0.05
634 may be achieved by populations naturally within a couple of generations, whereas larger changes,
635 such as those predicted for temperature derived variables (Figure 5), are unlikely to be achieved
636 naturally by long-living forest trees with long-generation times, even when they exhibit high levels of
637 standing genetic variation. This reinforces the importance of considering conservation strategies
638 such as assisted gene flow or assisted migration (Borrell et al., 2019).

639

640 **Concluding remarks**

641 This study provides insights into both the evolution and ecology of Asian white Birch, *B. platyphtylla*,
642 in China. We showed that *B. platyphylla* and *B. pendula* are better considered distinct species, as
643 their differentiation is clear at the genomic level. We identified a hybrid zone between *B.*
644 *platyphylla/B. pendula* in north-western China and three distinct lineages of *B. platyphylla* in eastern
645 and southern China distributed along a latitudinal gradient from the north-east to the south-west,
646 and. The three *B. platyphylla* lineages appear to have had a long period of isolation from one
647 another, followed by more recent gene flow. Our species distribution model shows that *B.*
648 *platyphylla* prefers mountainous environments with cool summer and moderate levels of
649 precipitations. Future prediction shows significant reduction on the suitable habitat of this species in
650 China under every future scenario tested, up to a third reduction compared to the current extent.
651 We identified signals of local adaptation throughout the genome of this species, suggesting that
652 climate adaptation is a highly polygenic mechanism. We estimated the degree of maladaptation to
653 inform conservation strategies. Despite the limitation of our method, we showed that rapidly rising
654 temperature, particularly in the summer months, poses a risk to this species and our environmental
655 niche model suggests that current habitats located in central China and in the north-eastern region
656 adjacent to the Changbai mountains will likely become unsuitable by 2080-2100, while current
657 habitats located at higher elevations, such as the south-western Tibetan plateau and the region
658 surrounding the Greater Khingan mountains range in the north-east will remain suitable. It is
659 possible that populations outside the predicted suitable environment may be able to adapt, however
660 climate, particularly temperature, might change to an extent that adaptation will not be possible and
661 assisted migration might be the only option possible to rescue some populations. The estimated
662 required changes in allelic frequencies at adaptive loci do not appear as concerning within the
663 predicted suitable environment, however with the current dataset, based on individuals' allelic
664 frequencies, it is difficult to estimate with confidence an allelic frequency change threshold after

665 which populations will fail to adapt on time naturally and will require assisted gene flow from donor
666 pre-adapted populations.

667

668 **Acknowledgements**

669 This work was funded by the National Natural Science Foundation of China (31770230 and
670 31600295) to Nian Wang. Gabriele Nocchi is on a PhD programme run jointly between Queen Mary
671 University of London and the Royal Botanic Gardens, Kew funded by the UK Government
672 Department for Environment, Food and Rural Affairs (Defra), in association with Action Oak.

673 We thank Stefan Götz, David Seide and the BioBam Bioinformatics team for kindly letting us use
674 their software OmicsBox for the functional annotation and functional enrichment analysis of the
675 identified adaptive genes. We thank Richard Nichols (Queen Mary University of London), Yiye Liang
676 and Baosheng Wang (South China Botanical Garden, Chinese Academy of Sciences) for their expert
677 advice on our analyses.

678 This research utilised Queen Mary's Apocrita HPC facility, supported by QMUL Research-IT:
679 http://doi.org/10.5281/zenodo.438045.

680

681 **References**

682 1. Aitken, S., Yeaman, S., Holliday, J., Wang, T., & Curtis-McLane, S. (2008). Adaptation,
683 migration or extirpation: climate change outcomes for tree populations. *Evolutionary*
684 *Applications*, *1*(1), 95-111.

685 2. Alberto, F., Aitken, S., Alía, R., González-Martínez, S., Hänninen, H., & Kremer, A. et al. (2013).
686 Potential for evolutionary responses to climate change – evidence from tree populations.
687 *Global Change Biology*, *19*(6), 1645-1661.

688 3. Ashburner, K., & McAllister, H. (2013). Botanical Magazine Monograph: The Genus Betula: A
689 Taxonomic Revision of Birches. Erscheinungsort nicht ermittelbar: Royal Botanic Gardens.

690 4. Ashburner, M., Ball, C., Blake, J., Botstein, D., Butler, H., & Cherry, J., … Sherlock, G. (2000).
691 Gene Ontology: tool for the unification of biology. *Nature Genetics*, *25*(1), 25-29.

5. Bai, W.-N., Wang, W.-T., & Zhang, D.-Y. (2015). Phylogeographic breaks within Asian butternuts indicate the existence of a phytogeographic divide in East Asia. *New Phytologist, 209*(4), 1757–1772.

6. Birks, H., & Willis, K. (2008). Alpines, trees, and refugia in Europe. *Plant Ecology & Diversity*, 1 (2), 147-160.

7. Bolger, A., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120.

8. Borrell, J., Zohren, J., Nichols, R., & Buggs, R. (2019). Genomic assessment of local adaptation in dwarf birch to inform assisted gene flow. *Evolutionary Applications*, *13*(1), 161-175.

9. Browne, L., Wright, J. W., Fitz-Gibbon, S., Gugger, P. F., & Sork, V. L. (2019). Adaptational lag to temperature in valley oak (Quercus lobata) can be mitigated by genome-informed assisted gene flow. *Proceedings of the National Academy of Sciences*, *116*(50), 25179–25185.

10. Cao, X., Herzschuh, U., Ni, J., Zhao, Y., & Böhmer, T. (2015). Spatial and temporal distributions of major tree taxa in eastern continental Asia during the last 22,000 years. *The Holocene*, *25*(1), 79-91.

11. Caye, K., Deist, T., Martins, H., Michel, O., & François, O. (2016). TESS3: fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources*, *16*(2), 540-548.

12. Caye, K., Jumentier, B., Lepeule, J., & François, O. (2019). LFMM2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies. *Molecular Biology and Evolution*, *36*(4), 852-860.

13. Chang C. C., Chow, CC, Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015) Second-generation PLINK: rising to the challenge of larger and ricer datasets. *GigaScience*, *4*(1)

14. Chen, S., Wang, Y., Yu, L., Zheng, T., Wang, S., & Yue, Z. et al. (2021). Genome sequence and evolution of *Betula platyphylla. Horticulture Research*, *8*(1).

15. Chen, T., & Lou, A. (2019). Phylogeography and paleodistribution models of a widespread birch (*Betula platyphylla* Suk.) across East Asia: Multiple refugia, multidirectional expansion, and heterogeneous genetic pattern. *Ecology And Evolution*, *9*(13), 7792-7807.

16. Conesa, A., & Götz, S. (2008). Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. *International Journal of Plant Genomics*, 1-12.

17. Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M., … Durbin, R. and 1000 Genomes Project Analysis Group. (2011). The Variant Call Format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.

18. Dauphin, B., Rellstab, C., Schmid, M., Zoller, S., Karger, D., & Brodbeck, S., Guillaume, F., & Gugerli, F. (2020). Genomic vulnerability to rapid climate warming in a tree species with a long generation time. *Global Change Biology*, 27(6), 1181-1195.

19. Davey, J., Hohenlohe, P., Etter, P., Boone, J., Catchen, J., & Blaxter, M. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, *12*(7), 499-510.

20. DePristo, M., Banks, E., Poplin, R., Garimella, K., Maguire, J., & Hartl, C., …, & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491-498.

21. Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for Ancient Admixture between Closely Related Populations. *Molecular Biology and Evolution*, *28*(8), 2239–2252.

22. Eckert, A., van Heerwaarden, J., Wegrzyn, J., Nelson, C., Ross-Ibarra, J., González-Martínez, S., & Neale, D. (2010). Patterns of population structure and environmental associations to aridity across the range of Loblolly Pine (*Pinus taeda* L., Pinaceae). Genetics, *185* (*3*), 969-982.

23. Ekblom, R., & Galindo, J. (2010). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, *107*(1), 1-15.

24. Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends In Ecology & Evolution*, *29*(1), 51-63.

25. Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V., & Foll, M. (2013). Robust Demographic Inference from Genomic and SNP Data. *Plos Genetics*, *9* (10), e1003905.

26. Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V., & Foll, M. (2013). Robust Demographic Inference from Genomic and SNP Data. *Plos Genetics*, *9* (10), e1003905.

27. Fan, D. M., Yue, J. P., Nie, Z. L., Li, Z. M., Comes, H. P., & Sun, H. (2013). Phylogeography of *Sophora davidii* (Leguminosae) across the 'Tanaka-Kaiyong Line', an important phytogeographic boundary in Southwest China. *Molecular Ecology*, *22*(16), 4270–4288.

28. Fisher, R. (1930). The genetical theory of natural selection. Oxford: Oxford University Press.

29. Frichot, E., Schoville, S., 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.

30. Gain, C., & François, O. (2021). LEA 3: Factor models in population genetics and ecological genomics with R. *Molecular Ecology Resources.*

31. Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H. Y., Hansen, N. F., Durand, E. Y., Malaspinas, A. S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prufer, K., Meyer, M., Burbano, H. A., . . . Paabo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science*, *328*(5979), 710–722. https://doi.org/10.1126/science.1188021

32. Guichoux, E., Garnier-Géré, P., Lagache, L., Lang, T., Boury, C., & Petit, R. (2012). Outlier loci highlight the direction of introgression in oaks. *Molecular Ecology*, *22*(2), 450-462.

33. Hewitt, G. (1999). Post-glacial re-colonization of European biota. *Biological Journal Of The Linnean Society*, 68(1-2), 87-112.

34. Hijmans, R. J., & Etten, J. (2012). raster: Geographic analysis and modelling with raster data.

35. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high-resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978.

36. Hoban, S., Kelley, J., Lotterhos, K., Antolin, M., Bradburd, G., & Lowry, D. et al. (2016). Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American Naturalist*, *188*(4), 379-397.

37. Joost, S., Bonin, A., Bruford, M. W., Després, L., Conord, C., Erhardt, G., & Taberlet, P. (2007). A spatial analysis method (SAM) to detect candidate loci for selection: Towards a landscape genomics approach to adaptation. *Molecular Ecology*, 16, 3955–3969.

38. Jordan, R., Hoffmann, A., Dillon, S., & Prober, S. (2017). Evidence of genomic adaptation to climate in Eucalyptus microcarpa: Implications for adaptive potential to projected climate change. *Molecular Ecology*, *26*(21), 6002-6020.

39. Jorhi, P., Eyre-Walker, A., Gutenkunst, R. N., Lohmueller, K. E., & Jensen, J. D. (2022). On the prospect of achieving accurate joint estimation of selection with population history. *Genome Biology and Evolution*, evac088.

40. Jump, A. S., Hunt, J. M., Martinez-Izquierdo, J. A., & Peñuelas, J. (2006). Natural selection and climate change: Temperature-linked spatial and temporal trends in gene frequency in Fagus sylvatica. *Molecular Ecology*, *15*, 3469–3480.

41. Jun-Wei, Y., Yang, Z., & Xiao-Juan, W. (2017). Phylogeographic breaks and the mechanisms of their formation in the Sino-Japanese floristic region. *Chinese Journal of Plant Ecology*, *41*(9), 1003–1019.

42. Keenan, R. (2015). Climate change impacts and adaptation in forest management: a review. *Annals of Forest Science*, *72*(2), 145-167.

43. Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, *15* (1), 356.

44. Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754-1760.

45. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., & Homer, N., …, & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079.

46. Liu, C., Newell, G., & White, M. (2016). On the selection of thresholds for predicting species occurrence with presence-only data. *Ecology and Evolution*, *6*, 337–348.

47. Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. Molecular Ecology, 23(9), 2178–2192.

48. Malinsky, M., Matschiner, M., & Svardal, H. (2020). Dsuite - Fast D -statistics and related admixture evidence from VCF files. *Molecular Ecology Resources.*

49. Manel, S., & Holderegger, R. (2013). Ten years of landscape genetics. *Trends In Ecology & Evolution*, *28*(10), 614-621.

50. Martins, H., Caye, K., Luu, K., Blum, M. G. B., & François, O. (2016). Identifying outlier loci in admixed and in continuous populations using ancestral population differentiation statistics. Molecular Ecology, 25(20), 5029–5042.

51. Masson-Delmotte, V., P. Zhai, A. Pirani, S.L., Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R., Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and Zhou, B. (2021). IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press. In Press.

52. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., & Kernytsky, A., …, & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analysing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297-1303.

53. Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, *76*(10), 5269–5273.

54. Nosil, P., Funk, D., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, *18*(3), 375-402.

55. Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modelling of species geographic distributions. *Ecol. Modell.*, *190*, 231–259.

56. Pina-Martins, F., Baptista, J., Pappas, G., & Paulo, O. (2018). New insights into adaptation and population structure of cork oak using genotyping by sequencing. *Global Change Biology*, *25*(1), 337-350.

57. Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841-842.

58. Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics*, *197* (2), 573-589.

59. Rellstab, C., Dauphin, B., & Exposito-Alonso, M. (2021). Prospects and limitations of genomic offset in conservation management. *Evolutionary Applications*, *14*(5), 1202-1212.

60. Rellstab, C., Gugerli, F., Eckert, A., Hancock, A., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, *24*(17), 4348-4370.

61. Rellstab, C., Zoller, S., Walthert, L., Lesur, I., Pluess, A., & Graf, R. et al. (2016). Signatures of local adaptation in candidate genes of oaks (*Quercus* spp.) with respect to present and future climatic conditions. *Molecular Ecology*, *25*(23), 5907-5924.

62. Salojärvi, J., Smolander, OP., Nieminen, K., Rajaraman, S., Safronov, O., Safdari, P., … Kangasjarvi, J. (2017). Genome sequencing and population genomic analyses provide insights into the adaptive landscape of silver birch. *Nature Genetics*, *49*, 904–912.

63. Sami, A., Arabia, S., Sarker, R., & Islam, T. (2021). Deciphering the role of helicases and translocases: A multifunctional gene family safeguarding plants from diverse environmental adversities. *Current Plant Biology*, *26*, 100204.

64. Savolainen, O., Pyhäjärvi, T., & Knürr, T. (2007). Gene Flow and Local Adaptation in Trees. *Annual Review of Ecology, Evolution, and Systematics*, *38*(1), 595-619.

65. Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, *46* (8), 919-925.

66. Sharma, T., Dreyer, I., & Riedelsberger, J. (2013). The role of K+ channels in uptake and redistribution of potassium in the model plant *Arabidopsis thaliana*. Frontiers in Plant Science, 4.

67. Sork, V., Aitken, S., Dyer, R., Eckert, A., Legendre, P., & Neale, D. (2013). Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes*, *9*(4), 901-911.

68. Storey, J. D., Bass, A. J., Dabney, A., & Robinson, D. (2015). qvalue: Q-value estimation for false discovery rate control. R package version 2.4.2. http://github.com/jdstorey/qvalue

69. Storey, J., & Tibshirani, R. (2003). Statistical significance for genome wide studies. *Proceedings Of the National Academy Of Sciences*, *100*(16), 9440-9445.

70. Stucki, S., Orozco-terWengel, P., Forester, B. R., Duruz, S., Colli, L., Masembe, C.…Consortium, N. (2017). High performance computation of landscape genomic models including local indicators of spatial association. *Molecular Ecology Resources*, *17*(5), 1072–1089.

71. Tsuda, Y., Semerikov, V., Sebastiani, F., Vendramin, G., & Lascoux, M. (2017). Multispecies genetic structure and hybridization in the *Betula* genus across Eurasia. *Molecular Ecology*, *26*(2), 589-605.

72. Wang, I., & Bradburd, G. (2014). Isolation by environment. *Molecular Ecology*, *23* (*23*), 5649-5662.

73. Wang, N., Kelly, L., McAllister, H., Zohren, J., & Buggs, R. (2021). Resolving phylogeny and polyploid parentage using genus-wide genome-wide sequence data from birch trees. *Molecular Phylogenetics and Evolution*, *160*, 107126.

74. Wang, N., McAllister, H., Bartlett, P., & Buggs, R. (2016). Molecular phylogeny and genome size evolution of the genus *Betula* (Betulaceae). *Annals of Botany*, *117*(6), 1023-1035.

75. Wang, N., Thomson, M., Bodles, W., Crawford, R., Hunt, H., & Featherstone, A., Pellicer, J., Buggs, R. J. A. (2013). Genome sequence of dwarf birch (*Betula nana*) and cross-species RAD markers. *Molecular Ecology*, *22*(11), 3098-3111.

76. Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*, 1358–1370.

77. Zdobnov, E., & Apweiler, R. (2001). InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, *17*(9), 847-848.

78. Zhang, C., Dong, S., Xu, J., He, W., & Yang, T. (2018). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*, *35*(10), 1786-1788.

79. Zhou, B., Shi, Y., Chen, X., Yuan, S., Liang, Y., & Wang, B. (2022). Linked selection, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence in Quercus dentata. *Journal of Systematics and Evolution*.

888 **Data accessibility**

889 Sequencing read data are archived in the NCBI through the BioProject accession PRJNA854976.

890

891 **Benefit sharing statement**

892 This research was led by scientists from the country providing genetic samples. As described above,

893 all data have been shared with the broader public via appropriate biological databases.

894

895 **Author contributions**

896 RB and NW conceived the project. NW and JD performed the sampling and extracted DNA. LY, JD,

897 and YG performed mapping and SNPs calling. GN and JW analysed the data and wrote the

898 manuscript. RB oversaw the project and helped write the manuscript.

899

900

901

902

903

904

905

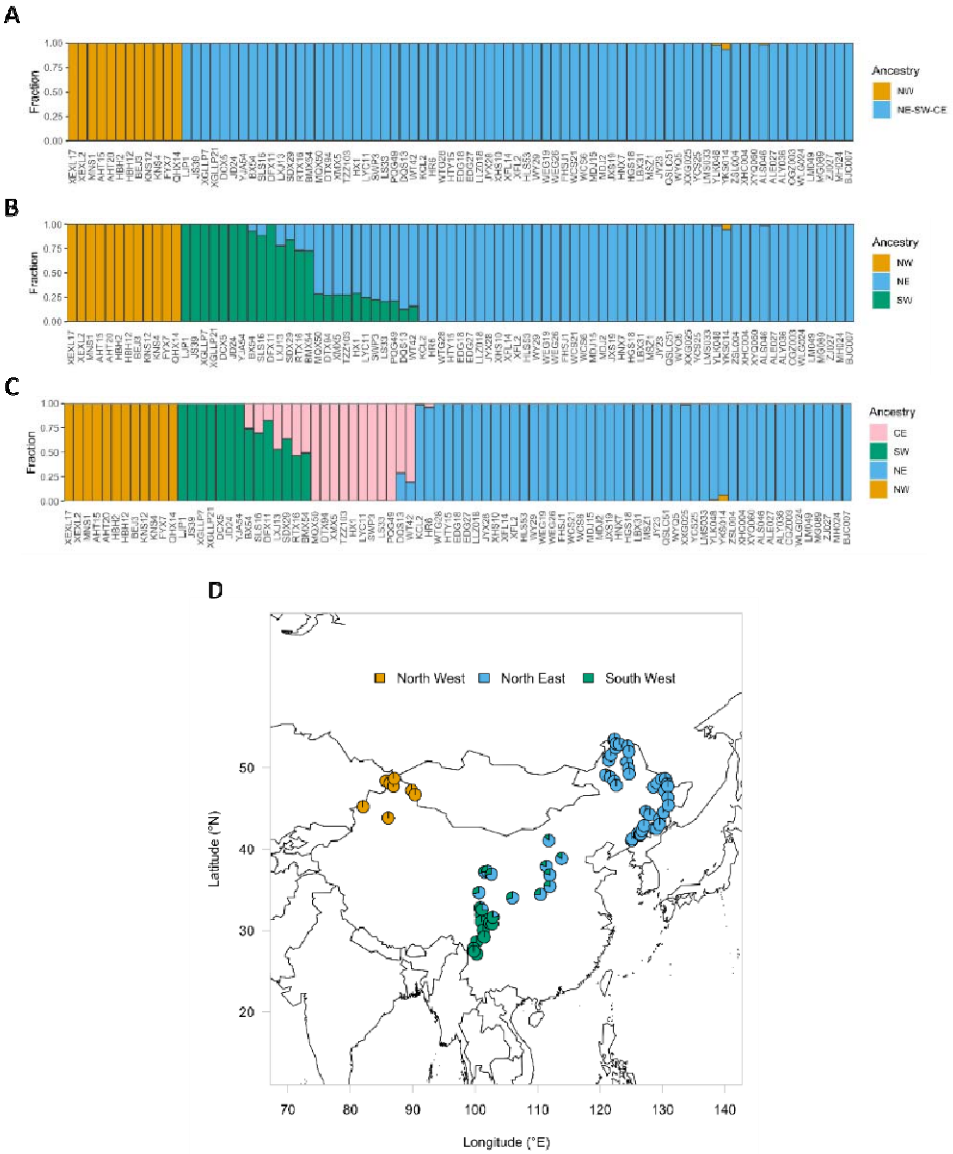
906

907

908 **Figures and Tables**

909 **Figure 1.** *fastSTRUCTURE* results for the China dataset, ordered by longitude. **A)** Bar plot of
910 *fastSTRUCTURE* results of 83 sampled birch individuals, based on 1,497,547 unlinked SNPs ($r^2 < 0.4$)
911 at K = 2, **B)** K=3 and **C)** K =4. **D)** Map of the individuals showing their genetic ancestry composition
912 according to *fastSTRUCTURE* at the ideal model complexity, K = 3.

913



914

915

916

917

918

919

Figure 2. *fastSTRUCTURE* results for the Eurasian dataset, ordered by longitude. The size of the pies is proportional to the number of samples. The labelled bars correspond to individuals sampled in this study, while the unlabelled bars are the samples added from Salojarvi et al. (2017) **A)** Bar plot of *fastSTRUCTURE* results of 162 *B. pendula* and *B. platyphylla* individuals, based on 278,717 unlinked SNPs ($r^2 < 0.4$) at K = 2, **B)** K=3 and **C)** K =4. **D)** Map of the individuals showing their genetic ancestry composition according to *fastSTRUCTURE* at K = 4.

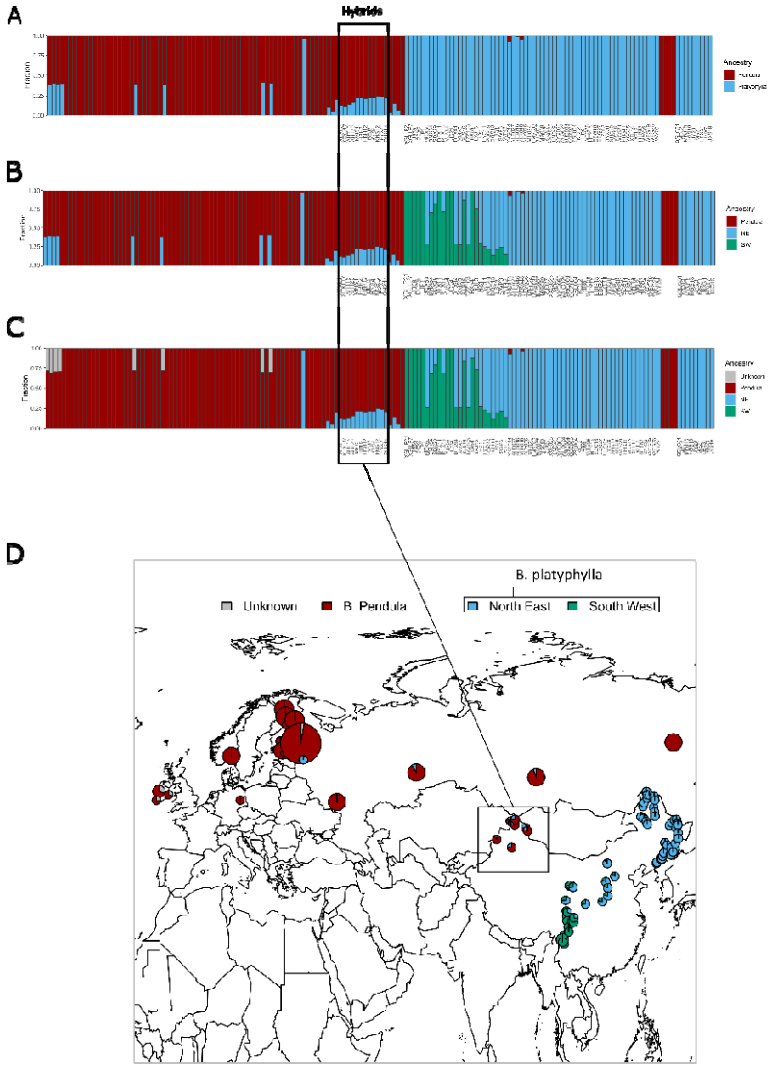


Figure 3. *fastSTRUCTURE*, *MSMC* and *fastsimcoal* results for Chinese *B. platyphylla,* ordered by longitude. **A)** Bar plot of *fastSTRUCTURE* results of 71 *B. platyphylla* individuals, based on 1,387,994 unlinked SNPs ($r^2 < 0.4$) at K = 2, and **B)** K=3. **C)** Map of the sampled individuals showing their genetic ancestry composition according to *fastSTRUCTURE* at K = 3. **D)** A demographic scenario of the NE, CE and SW lineages modelled using fastsimcoal2. Shown in light grey are the ancestral populations. The arrows show the per generation migration rate (m) between the three lineages. **E)** Historical effective population size of the three populations estimated by MSMC. Solid lines and shading represent medians and ±standard deviation calculated across pairs of haplotypes, respectively. **F)** The results of "ABBA-BABA" tests. The red line denotes topologies with significant signals of introgression, which are indicated by a mean Z scores > 3.09. EU and platy represent *B. pendula* and *B. platyphylla*, respectively.

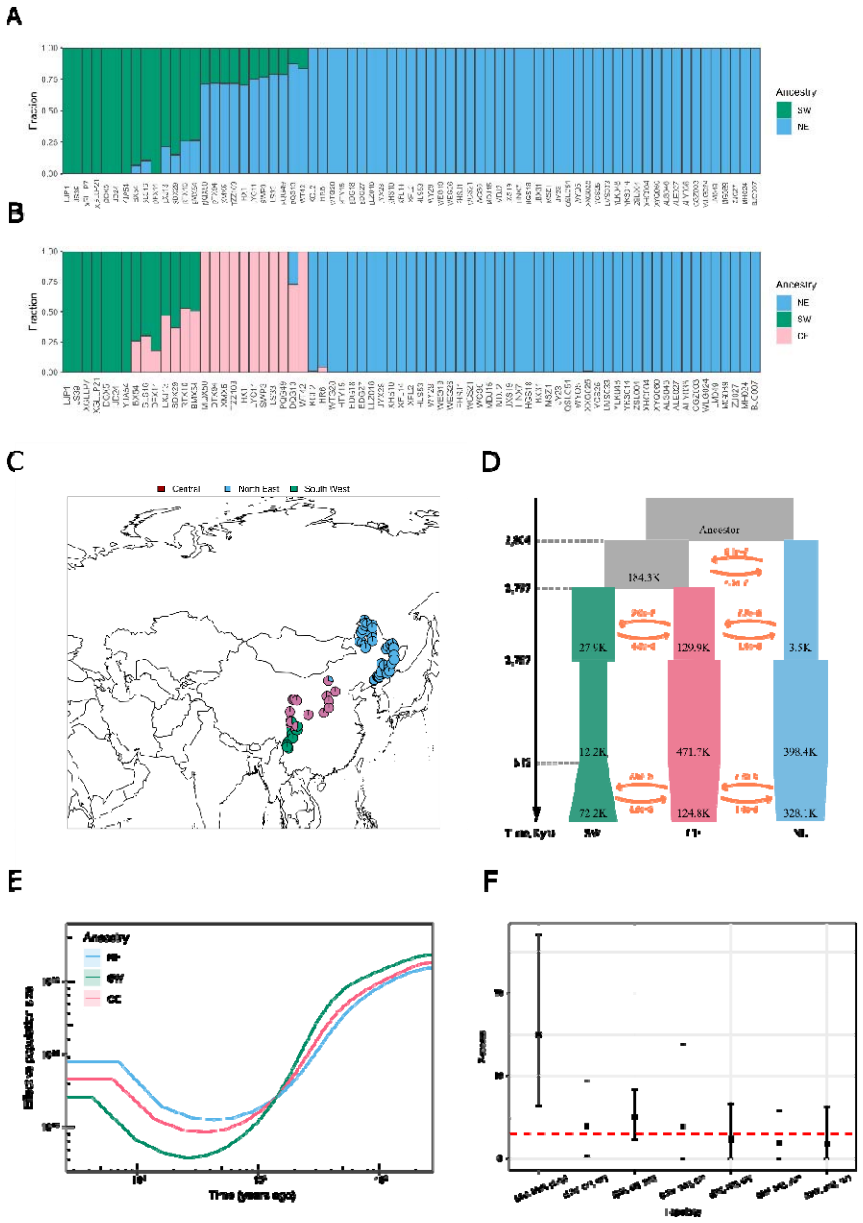


Figure 4. Environmental niche model (ENM) of *B. platyphylla* in China. Red points represent the individuals sampled in this study included in the model, blue points represent observations downloaded from the GBIF (http://www.gbif.org). **A)** Present time ENM showing the habitat suitability index (HSI) throughout China using a coloured scale from 0 (white) to 1 (green). **B)** Binomial representation of the HSI model, using the maximum training sensitivity plus specificity threshold of 0.2. **C)** Binomial HSI projections (> 0.2) of *B. platyphylla* under four different future climate scenarios (columns: ssp126, ssp245, ssp370 and ssp585) at two different time points (rows: 2041-2060 and 2080-2100). The percentage in each plot shows the decrease in suitable environment area compared with the current habitat.

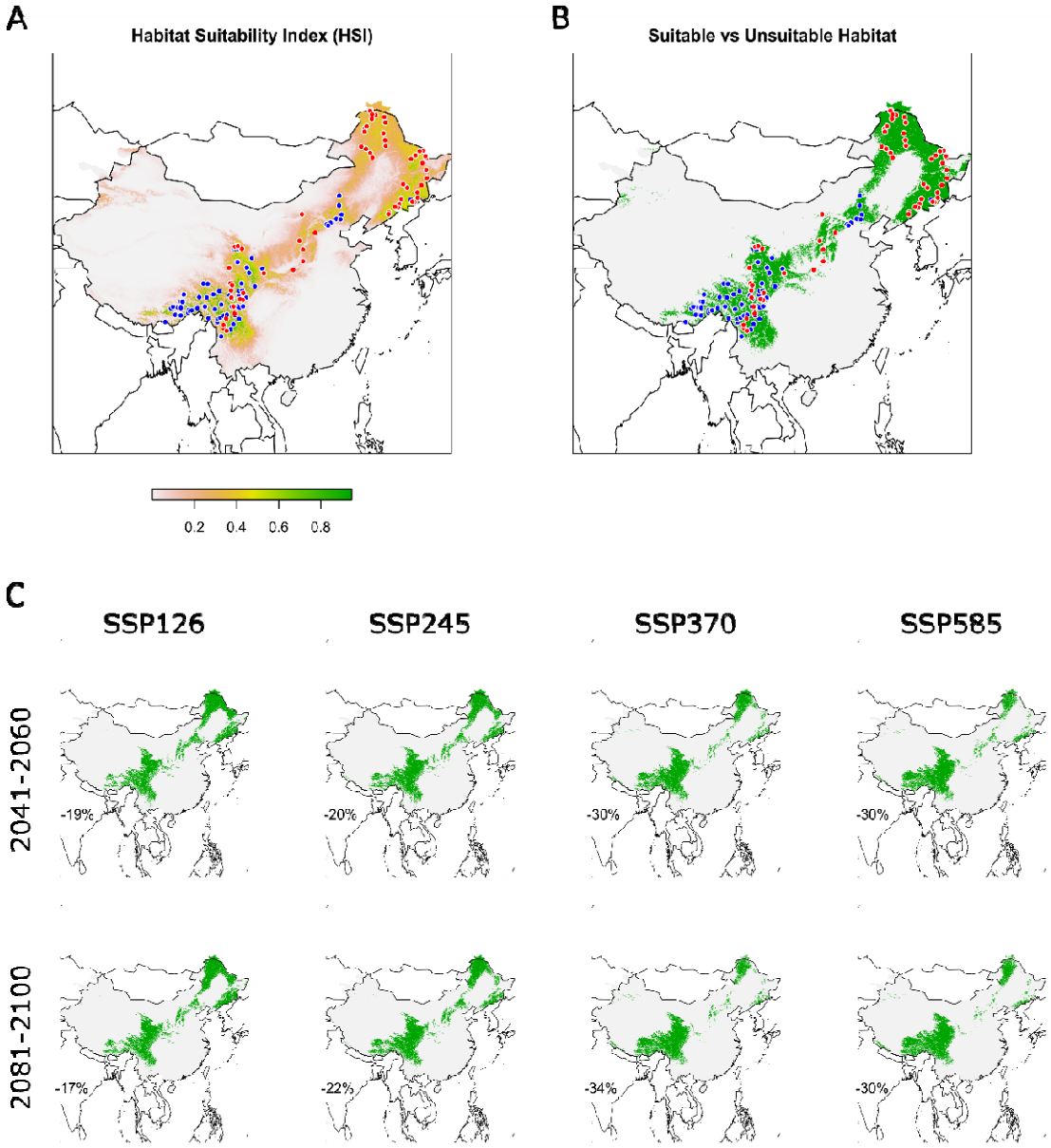


Figure 5. Risk of non-adaptedness (RONA) for the 71 *B. platyphylla* individuals**.** RONA calculated independently for each of the seven climatic variables projected to change in the future. The future climate profile ssp370 at 2080-2100 was used for RONA calculation. AD: admixed individuals. AMT: annual mean temperature. MDR: mean diurnal range. ISO: isothermality. MTWQ: mean temperature of wettest quarter. AP: annual precipitation. PS: precipitation seasonality. PDQ: precipitation of driest quarter.

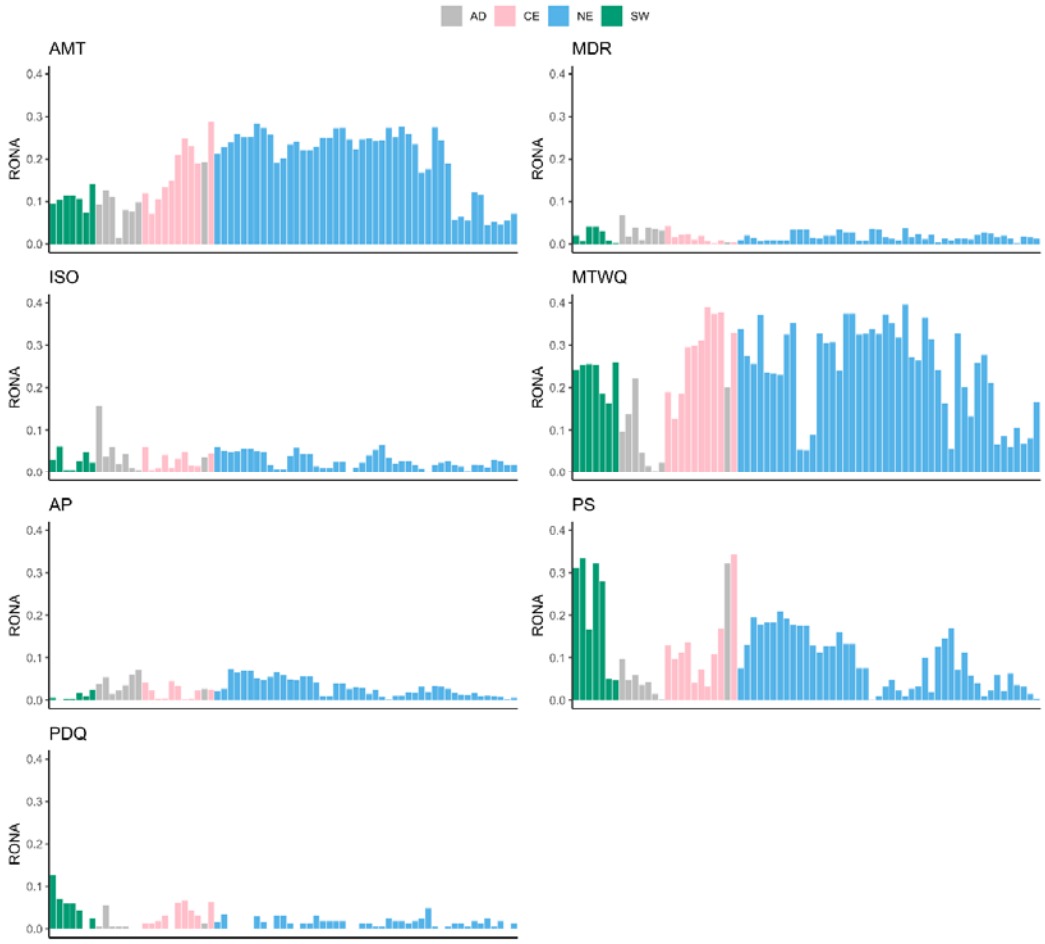
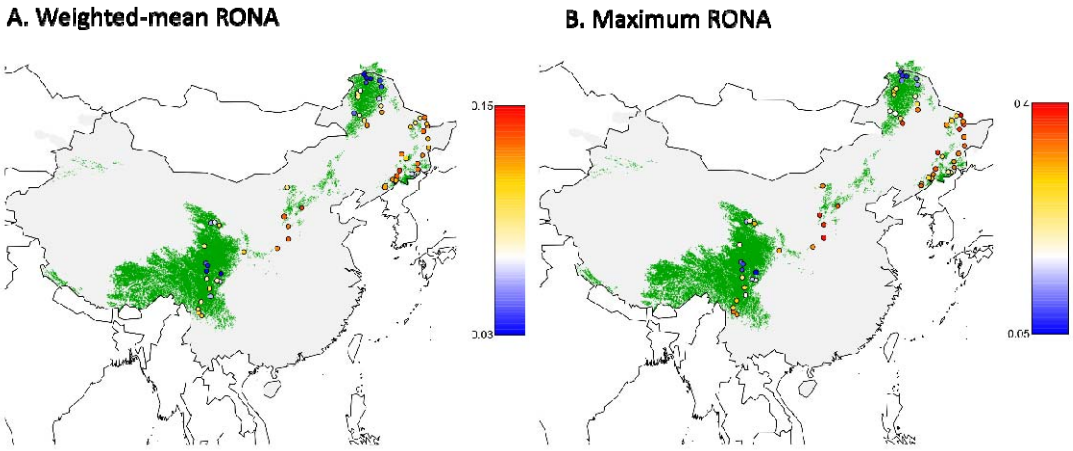



Figure 6. Weighted-mean RONA and MAX RONA. **A)** Map showing weighted mean RONA for the 71 *B. platyphylla* individuals, out of the seven environmental variables included in the analysis. Each variable RONA was given a weight equivalent to its percent contribution in the ENM. **B)** Map showing maximum RONA per individual, out of the seven environmental variables tested. Green shading represents suitable environment in 2080-2100 under profile ssp370 according to our ENM.