Multiple glacial refugia across northern and southern China and unexpected patterns of spatial genetic diversity in *Betula albosinensis*: a widespread temperate deciduous tree species

Lu Liu[1#], Andrew V. Gougherty[2#], Junyi Ding[1], Kun Li[1], Wenting Wang[3], Luwei Wang[1], Feifei Wang[1], Nian Wang[1,4,5*]

[1]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.

[2]Department of Botany, University of British Columbia, Vancouver, B.C., Canada

[3]School of Mathematics and Computer Science, Northwest Minzu University, Lanzhou 730030, China

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

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

[#]contributed equally to this work

[*]Author for correspondence: Nian Wang

Email: nian.wang@sdau.edu.cn

Abstract

The central-marginal hypothesis (CMH) predicts high genetic diversity at the species' geographic centre and low genetic diversity at the species' geographic margins. However, most studies examining the CMH have neglected potential effect of past climate. Here, we test six hypotheses, representing effects of past climate and contemporary range position, for their ability to explain the spatial patterns of genetic diversity in 37 populations of *Betula albosinensis*. Ecological niche modelling (ENM) revealed large and continuous suitable habitats in north, southwest and southeast China during the last glacial maximum (LGM) but a contraction of suitable habitats since the LGM. Pollen records further confirmed the existence of multiple refugia in north and south China. The spatial pattern of genetic diversity (i.e., expected heterozygosity, gene diversity and allele richness) were best explained by distance to the southern edge and distance to the range edge but also showed longitudinal and latitudinal gradients. Hypotheses accounting for the effects of climate (climatic suitability, climatic stability and climatic variability) had comparatively little support. Our findings show partial support for the CMH and illustrates that the existence of multiple LGM refugia, and suggests species abundance and past species movement play a role in shaping genetic diversity across species' ranges.

Keywords birch, ecological niche modelling, microsatellite markers, pollen record, refugia

Introduction

Developing an understanding of the factors shaping the spatial pattern of genetic diversity in species ranges will be important in predicting the response of populations to future climate change and informing population conservation. Populations' position within a species' range has been thought to impact the distribution of genetic diversity (Eckert, et al. 2008). For example, one of the most commonly tested hypotheses posits that populations located at the centre of the range tend to have a higher level of genetic diversity than populations at the geographic margin (Eckert, et al. 2008, Micheletti and Storfer 2015). This has been referred as the "central-marginal hypothesis" (CMH), and is thought to result from central populations having higher population abundance, larger effective population size, stronger gene flow and being nearer the ecological optimum compared to marginal populations (Sagarin and Gaines 2002). As a consequence, central populations are expected to have higher genetic diversity within populations and lower genetic differentiation among populations compared with marginal populations (Eckert, et al. 2008, Swaegers, et al. 2013).

However, the "central-marginal hypothesis" is often violated as the underlying pattern thought to result in the CMH, i.e. high population abundance in the centre of species' range is not consistently found (Dallas, et al. 2017, Sagarin and Gaines 2002). Historical climate and species' response to the changing climate may also impact the spatial pattern of genetic diversity. For example, temperature shifts during the glacial and inter-glacial cycles, especially the last glacial maximum (LGM), often required

forest trees to retreat southward into geographically isolated refugia (Hewitt 1999, 2004). As a consequence of this southern retreat and eventual northward recolonisation, many species exhibit a latitudinal decrease in genetic diversity (Petit, et al. 1997); although cryptic refugia and admixture zones are known to obscure simple latitudinal gradients (Petit, et al. 2003). Species' dispersal ability may also impact the distribution of genetic diversity. For example, species with high dispersal ability sometimes leave a more obvious genetic gradient than species with low dispersal ability (Wang, et al. 2016, Ye, et al. 2019).

The LGM is thought to have had a more minor impact on the distribution of forest trees in East Asia compared to other regions in the northern hemisphere (Qian and Ricklefs 2001, Qiu, et al. 2011). Based on pollen fossil records, it has been proposed that most temperate forest trees in East Asia retreated to between 25 and 30 N during the LGM (Qian and Ricklefs 2001). However, an increasing number of phylogeographic studies indicate the existence of a single refugium in northeast China (NEC) or multiple isolated refugia in NEC and/or north China (NC) for temperate cold-tolerant trees (Hou, et al. 2018, Hu, et al. 2008, Liu, et al. 2012, Zeng, et al. 2015, Zhang, et al. 2005). Several studies indicate an expansion from a single refugium into northeastern China, resulting in a decrease in genetic diversity with increasing latitude (Hu, et al. 2008). Other studies favor multiple isolated refugia in north China, despite a decrease in genetic diversity (Zeng, et al. 2015, Zhang, et al. 2005), or a mixed pattern of genetic diversity, with populations in NEC showing a latitudinal decrease in genetic diversity and a latitudinal increase in genetic diversity in south China (Liu, et

al. 2012).

Hence, understanding the drivers of genetic diversity in species ranges requires incorporating information about past climate and demographic history in the frame of CMH. However, to date, only a small number of studies consider past climate (Gougherty, et al. 2020, Jin, et al. 2020). By doing so, some studies have shown that the effects of past climate are relatively minor, while others indicate that historical factors strongly influences the spatial patterns of genetic variation. In addition, many studies do not disentangle historical and contemporary range position and do not account for range-wide spatial auto-correlation.

In this study, we tested whether historical processes affect spatial patterns of genetic diversity in the context of the CMH in China (Wei, et al. 2016). Using *Betula albosinensis* as our study species, we compared six hypotheses for their ability to describe the patterns of genetic diversity across 37 populations. *Betula* is an ecologically important genus that consists of approximately 65 species and subspecies (Ashburner and McAllister 2016, Wang, et al. 2016), widely distributed across the Northern Hemisphere. *Betula* species are wind-pollinated and self-incompatible, often resulting in a high level of genetic diversity within populations (Ashburner and McAllister 2016). Some *Betula* species are pioneers, colonising new habitats and providing shelter for other trees.

Our focal species, *B. albosinensis*, is a deciduous broad-leaved temperate species,

growing in mountains at an altitude between 1800~3800m according to our field observations. In China, it extends from northwestern Yunnan province to northern Hebei province, between 25° and 40° latitude and between 99° and 116° longitude. *Betula albosinensis* is a pioneer species and can grow up to 35 meters. Its regeneration depends on habitat disturbances as evidenced by the observations that seedlings are only found in open habitats caused by tree gaps (Guo, et al. 2019).

Here, we used 16 microsatellite markers to investigate the spatial distribution of genetic diversity of *B. albosinensis*. We used ENM and pollen records to robustly infer its suitable habitats since the LGM, and based on these paleo reconstructions of its distribution, we further explored factors impacting the distribution of its population genetic diversity. The specific questions we sought to address are: (1) Does *B. albosinensis* have multiple LGM refugia in China? (2) Does *B. albosinensis* form different genetic clusters? (3) What is the spatial distribution of genetic diversity? (4) What factors impact the spatial distribution of genetic diversity?

Materials and Methods

Field sampling

Betula albosinensis samples were collected from natural populations over four years (2016-2019), covering nearly its entire distribution in China. Within each population,

131 samples were chosen at random and separated by at least 20 meters. A GPS system
132 (UniStrong) was used to record sampling locations. Herbarium specimens were
133 collected from most individuals and, for a subset of individuals where twigs were out
134 of our reach, cambial tissue was collected and stored in coffee bags and dried using
135 silica gel. A total of 815 individuals were collected from 37 populations, with 3-56
136 individuals sampled from each population (Table S1).
137
138 **Species distribution model**
139
140 To predict the potential distribution of *B. albosinensis*, we used occurrence records
141 from our own field work recorded with a GPS system (UniStrong), and collected
142 additional occurrence records from the literature published since the year 2000. If *B.*
143 *albosinensis* was recorded in a particular certain region, but lacked a geographic
144 coordinate, a random point was selected to represent its occurrence within the region.
145 To avoid spatial autocorrelation due to geographic aggregation, only one point was
146 kept every 5 km. A total of 264 points data were obtained and after filtering, 132
147 records were retained for ecological niche modelling.
148 Nineteen bioclimatic variables were downloaded from WorldClim
149 (http://www.worldclim.org) for the four periods: the Last Glacial Maximum (LGM),
150 the Middle Holocene (MID), Present (1970-2000) and Future (2050-2070) (Hijmans,
151 et al. 2005). Present climate variables are derived from monthly mean precipitation
152 and temperature data from the World Meteorological Station from 1970 to 2000 (Fick

153 and Hijmans 2017). Simulated climate data were selected for the other three periods
154 under the Community Climate System Model (CCSM4) (Gent, et al. 2011), in which
155 the data of Representative Concentration Pathways 85 (RCP 85) were selected for
156 FUTURE climate variables. The current and past distribution of *B. albosinensis* were
157 estimated based on an ensemble species model, performed using the R package
158 "BiodiversityR" (Kindt 2018). We selected six bioclimatic variables that lacked strong
159 correlation (<0.75), identified using "ENMTools" (Warren, et al. 2010), to include in
160 subsequent analyses: BIO01 (annual mean temperature), BIO03 (isothermality),
161 BIO07 (temperature annual range), BIO13 (precipitation of wettest month), BIO14
162 (precipitation of driest month) and BIO15 (precipitation seasonality).
163
164 Maximum entropy (MAXENT), generalized boosted regression modeling (GBM),
165 random forest (RF), generalized linear models (GLM) and support vector machines
166 (SVM) were selected for niche model integration simulation test, setting five
167 cross-validations using "BiodiversityR" (Kindt 2018). The five algorithms included in
168 the integrated model had equivalent weights (0.20) for each model. Models were
169 evaluated by splitting data into training and testing datasets with 80% of the data
170 being used to train and 20% to test the models. The true skill statistic (TSS) and the
171 area under the receiver operating characteristics (ROC) curve were used to assess the
172 performance of the models. TSS scores range from −1 to 1, where +1 indicates a
173 perfect ability to distinguish suitable from unsuitable habitat, while values of zero (or
174 less) indicate a performance no better than random. For the ensemble modelling, only

175 those models with a TSS value greater than 0.85 were considered.

176

177 **Pollen records**

178

179 To infer the past distribution of *B. albosinensis* in China, we collected pollen records
180 of *Betula* species from the published literature. For most pollen cores, paleobotanists
181 identified pollen only to the genus level; however, *Betula* pollen in this region is
182 likely *B. albosinensis* as it is currently the dominant *Betula* species. We mapped these
183 pollen sites using coordinates given in the literature. We grouped pollen records into
184 two time periods to coincide with projections from the distribution model: the LGM
185 (22Ka-19Ka) and the Early to Middle Holocene (19Ka-4Ka). The detailed pollen
186 records are listed in table S2.

187

188 **DNA extraction and microsatellite genotyping**

189

190 Genomic DNA was extracted from cambial tissue using a previously modified 2×
191 CTAB (cetyltrimethylammoniumbromide) protocol (Wang, et al. 2013). The quality
192 of genomic DNA was assessed with 1.0 % agarose gels and then was diluted to a
193 concentration of ~10 ng/ul for microsatellite genotyping. Sixteen microsatellite loci
194 developed from closely-related species (Kulju, et al. 2004, Truong, et al. 2005, Tsuda,
195 et al. 2008, Wu, et al. 2002) were used to genotype our samples. The 5' terminus of
196 the forward primers was labeled with FAM, HEX or TAM fluorescent probes. Each

197 microsatellite locus was amplified individually prior to being artificially combined
198 into four multiplexes. In order to avoid errors caused by size overlapping, loci with
199 significant length differences were labeled using the same dye. PCR procedures were
200 provided in supplementary data. PCR products were delivered to Personal (Shanghai)
201 for microsatellite genotyping. Microsatellite alleles were scored using the software
202 GENEMARKER 2.4.0 (Softgenetics) and checked manually. Individuals with more
203 than three missing loci were excluded, resulting in 815 individuals in the final dataset.
204
205 **Microsatellite data analyses**
206
207 Microsatellite data of *B. albosinensis* was analyzed in STRUCTURE 2.3.4 (Pritchard,
208 et al. 2000) to identify the most likely number of genetic clusters (K) and estimate
209 genetic admixture. As *B. albosinensis* is a tetraploid species, we set ploidy to four in
210 STRUCTURE. Ten replicates of the STRUCTURE analysis were performed with
211 1,000,000 iterations and a burn-in of 100,000 for each run at each value of K from 1
212 to 6. We used the admixture model, with an assumption of correlated allele
213 frequencies among populations. Individuals were assigned to clusters based on the
214 highest membership coefficient averaged over the ten independent runs. The number
215 of biologically-meaningful genetic clusters was estimated using the “Evanno test”
216 (Evanno, et al. 2005) and the program Structure Harvester (Earl and vonHoldt 2011).
217 Gene diversity (Nei 1987) and allelic richness (El Mousadik and Petit 1996) were
218 calculated in the software FSTAT 2.9.4 (Goudet 1995). The tetraploid genotypes were

219 treated as two diploid individuals as described by Tsuda et al. (2017) and Hu et al.
220 (2019). A principal coordinate analysis (PCoA) was also performed on microsatellite
221 data of *B. albosinensis* using POLYSAT (Clark and Jasieniuk 2011) implemented in R,
222 ver. 4.0.2 (R Core Team 2020), based on pairwise genetic distances calculated
223 according to Bruvo et al. (2004).
224
225 **Geographic and climatic centrality**
226
227 We used eight variables as predictors of gene diversity (*Gd*), allele richness (*Ar*) and
228 expected heterozygosity (*He*). These include: (a) distance from the geographic range
229 centre (geoCentre), (b) distance from the southern range edge (southernEdge), (c)
230 distance from the range edge (geoEdge), (d) current climatic suitability, (e) climatic
231 distance from the climatic niche centroid (climDist), (f) climatic stability, (g) climatic
232 variability since the LGM, and (h) population structure/genetic admixture. In order to
233 estimate the distribution of *B. albosinensis*, we generated an alpha hull (Rodríguez
234 and Pateiro 2010) around the 264 occurrence records representing the minimum
235 convex polygons (Burgman and Fox 2003) and calculated the distance between each
236 population and the nearest edge. To calculate distance from the geographic range
237 centre, we located the centroid of the alpha hull using "rgeos" (Bivand and Rundel
238 2018) and then calculated the geographic distance between each population and the
239 centroid (Dallas, et al. 2017, Lira-Noriega and Manthey 2014). The alpha hull was
240 also used to calculate population distance from the southern edge and the distance of

each population from the nearest geographic edge (Gougherty, et al. 2020). If populations were located exactly on the geographic boundary or the southern edge, we set the distance to one kilometer.

We calculated the climatic niche centroid by averaging values of the six climatic variables of the 264 coordinate points used for ecological niche modelling. We used Mahalanobis distance as an index to measure the climatic distance between each population and the climatic niche centroid using "statmatch" (D'Orazio 2015). We calculated climatic stability as the sum of suitability of LGM, MID and present (Ortego, et al. 2015, Yannic, et al. 2013); and climatic variability as the standard deviation of suitability of the three time scales (Gougherty, et al. 2020).

In addition to effects of geographic/climatic centrality and past climate, we also tested for effects of population structure and admixture among genetic clusters on *He*, *Gd* and *Ar*. Using admixture proportions, we calculated a population-level index of admixture. To do so, first we averaged admixture proportions across individuals within populations. We defined populations with genetic admixture between 0.4 and 0.6 as admixed.

We measured the spatial autocorrelation of *He*, *Gd* and *Ar* using Moran's Index (I), with −1.0 and 1.0 indicating perfect dispersion and perfect clustering, respectively. Correlograms of Moran's I for *He*, *Gd* and *Ar* were estimated in 200 km increments.

Significance was determined for both the correlograms and global statistic by comparing the observed statistic to 999 random permutations following Gougherty et al. (2020). Using the eight variables, we compared statistical support for models representing six hypotheses (Table 1). For each hypothesis, we created conditional autoregressive (CAR) models to account for the potential effects of spatial autocorrelation in genetic diversity. CAR models use a weighted estimate of the response variable at neighbouring locations, in addition to the explanatory variables, to parameterise the models (Lichstein, et al. 2002). Neighbourhoods were defined as all populations within 600 km of one another as this distance ensured each population had at least one neighbour and was the approximate maximum distance of continuous positive spatial autocorrelation. Models were compared using Nagelkerke R2, Akaike information criterion (AIC) and Akaike weights (Wagenmakers and Farrell 2004). Each explanatory variable was scaled to a mean of 0 and a standard deviation of 1, to facilitate the comparison of coefficient estimates (Schielzeth 2010).

Results

Ecological niche modelling

A final ensemble model was created by incorporating weighted runs from the MAXENT, GBM, RF, GLM and SVM models, with TSS ranging from 0.89 to 0.92 (TSS average = 0.90) and AUC ranging from 0.96 to 0.98 (AUC average = 0.97).

285 The ensemble model predicts high habitat suitability during the LGM in the
286 Qinling-Daba Mountains, northwestern Yunnan and southeastern China (Fig. 1A).
287 During the Holocene, suitable habitat for *B. albosinensis* expanded from the Qinling
288 Mountains into some parts of Shaanxi province and contracted from southeastern
289 China (Fig. 1B). Suitable habitat for *B. albosinensis* seems to have been stable from
290 the Holocene to the present (Fig. 1BC) and is predicted to shift westward in the future
291 (Fig. 1D).
292
293 **Pollen records**
294
295 During the LGM (22 to 19 Ka), eleven pollen records of *Betula* species existed in
296 southwest, central and north China (Fig. 1A). However, pollen records of *Betula*
297 species became more widespread since the beginning of the Holocene (19 to 4 Ka)
298 and appeared in regions representing the current distribution of *B. albosinensis* (Fig.
299 1B).
300
301 **Genetic structure and admixture**
302
303 Our STRUCTURE analyses identified two clusters (K = 2) as the optimal grouping,
304 according to the ΔK criterion (Fig. S1A). Cluster I, termed hereafter the northern
305 cluster, included populations from the Qinling Mountains and the north (Fig. 2AB).
306 Cluster II, termed the southern cluster hereafter, included several populations from

307 northwestern Yunnan (DQ, JS, LJA, LJS, LP and WX) and southern Sichuan (DC, JD,
308 JLX, LFG, SDX and YJA) (Fig. 2AB). Seven populations (RTX, BMX, MEK, BX,
309 EBY and SLS) located mainly in western Sichuan province showed roughly equal
310 genetic admixture between the northern and southern clusters (Fig. 2AB). The
311 northern and southern clusters were geographically separated by the Sichuan Basin.
312 At higher K values (three to six), populations located at the southern margin in
313 northwest Yunnan provinces (LP, LJS, LJA, JS, WZ and DQ) and on the northwest
314 periphery in Qinghai and Gansu provinces (DTX and XMX) remain genetically
315 distinct (Fig. S1B).
316
317 We treated individuals with genetic admixture greater than 0.9 from the northern
318 cluster as “northern individuals” and individuals with genetic admixture greater than
319 0.9 from the southern cluster as “southern individuals”. We treated individuals with
320 genetic admixture in between as admixed. Principal coordinate analysis (PCoA) based
321 on Bruvo’s genetic distances among all samples showed that the “northern
322 individuals” can be separated from the “southern individuals” by PC1 (Fig. 3).
323 Admixed individuals bridged the gap between the “northern individuals” and
324 “southern individuals” and overlapped substantially with individuals from the two
325 clusters. PC1, PC2 and PC3 could not distinguish admixed individuals from
326 individuals from either the northern cluster or the southern cluster. PC1, PC2 and PC3
327 explained 7.0%, 3.8% and 3.4% of the total variation, respectively (Fig. 3).
328

329 **Spatial patterns of genetic diversity**

330 *Gd*, *Ar* and *He* ranged from 0.72 to 0.83, from 4.04 to 5.07 and from 0.69 to 0.82,
331 respectively. Moran's *I* showed moderate but significant spatial correlations for *He* (I
332 = 0.19, $P < 0.001$), *Gd* (I = 0.22, $P < 0.001$) and *Ar* (I = 0.32, $P < 0.001$) in the 37
333 populations, indicating similar levels of genetic diversity among adjacent populations.
334 Correlograms of Moran's *I* for the three metrics of genetic diversity were positively
335 autocorrelated up to ~200 km, and negatively correlated at 1500 and 2000 km (Fig.
336 4).

337

338 In general, populations with high *He*, *Gd* and *Ar* tend to be situated in the northern
339 portion of the range and the geographic centre whereas populations with low *He*, *Gd*
340 and *Ar* tend to be located near the southern margin. *He*, *Gd* and *Ar* were positively
341 correlated with both latitude and longitude (Fig. 5). Moreover, we found that climatic
342 distance from the climatic niche centroid was significantly positively correlated (r =
343 0.49, $P < 0.01$) with distance from the geographic range centre and significantly
344 negatively correlated with distance from the range edge (r = - 0.64, $P < 0.01$),
345 indicating that the geographic centre substantially overlapped with the climatic niche
346 centroid.

347

348 **Spatial models of genetic diversity**

349

350 The best performing model for *Gd*, *He* and *Ar* included distance from the range edge

and distance from the southern edge. This model had the highest Akaike weights and highest Nagelkerke R2 (Table 1) for each of the diversity metrics. The spatial models incorporating past climate (climate stability, variability, current suitability and climate distance) tended to have less support, and had Akaike weights near zero (Table 1).

Discussion

Range dynamics and its implications

In our study, three lines of independent evidence jointly point towards multiple refugia and cryptic refugia throughout the current distribution of *B. albosinensis*. First, ecological niche models (ENM) predict large and continuous suitable habitats for *B. albosinensis* situated in north, south and southeast China during LGM. Second, the presence of *Betula* pollen dated to the LGM in north and south China further confirms the existence of *Betula* species in these regions. Third, the high genetic diversity within populations in northern China suggests the existence of refugia near these populations. Interestingly, although the ENM does not predict high suitability at LGM at the northern range margin of *B. albosinensis*, high genetic diversity within northern populations (ie., XLA, BA, PQG, LS and SWP) supports the existence of cryptic LGM refugia in this region. This finding, together with several previous studies (Tian, et al. 2009, Zeng, et al. 2015), refutes the hypothesis that temperate forests migrated to the south (25–30 N) during the LGM (Cao, et al. 2014, Harrison, et al. 2001).

373 Instead, our findings favor the "multiple refugia hypothesis" that some temperate trees
374 had multiple refugia in north China (Chen, et al. 2008, Hao, et al. 2018, Wang, et al.
375 2016).
376
377 Unexpectedly, our ENM shows that the suitable habitats for *B. albosinensis* are
378 comparable or even larger than its current distribution, indicating that the LGM may
379 have had little effect on the distribution of *B. albosinensis*. This is possibly due to
380 some unusual characteristics of this species. First, *B. albosinensis* is cold-tolerant and
381 has a broad distribution across north, central and south China. Second, *B. albosinensis*
382 populations occupy a broad altitudinal range between 1800 and 2800m in the Qinling
383 Mountains according to our own field observations. Together, these could indicate
384 that *B. albosinensis* responded to the glacial and interglacial cycles by shifting its
385 range altitudinally, rather than latitudinally. In addition, *B. albosinensis*, like some
386 other *Betula* species, is wind-pollinated and produces a large number of tiny winged
387 seeds (Ashburner and McAllister 2016). This would help *B. albosinensis* disperse
388 long distances and occupy open habitats. Moreover, regeneration of *B. albosinensis*
389 depends on disturbance. Previous studies of *B. albosinensis* in the Qinling Mountains
390 show that its seedlings only grow in open habitats caused by tree gaps (Guo, et al.
391 2019). These characteristics may have enabled *B. albosinensis* to not only tolerate
392 LGM but may also have allowed it to colonise unglaciated habitats quickly.
393
394 Surprisingly, our ENM reveals large suitable habitats for *B. albosinensis* in eastern

China during the LGM and disappearance of suitable habitats since the Holocene. *Betula albosinensis* is presently absent from eastern China and *B. luminifera* is the dominant species there. If *B. albosinensis* historically existed in eastern China, one possible scenario for its disappearance is due to its hybridisation with *B. luminifera*, given the extensive hybridisation between *Betula* species documented elsewhere (Anamthawat-Jónsson and Thórsson 2003, Bona, et al. 2018, Eidesen, et al. 2015, Tsuda, et al. 2017, Wang, et al. 2014). Further research characterizing patterns of the genetic admixture between the two species may help to understand this.

Distinct genetic clusters

Both PCoA and STRUCTURE grouped *B. albosinensis* into two genetic clusters: a northern cluster from the Qinling-Daba Mountains and the regions to the north and northwest, and a southern cluster from northwestern Yunnan and southern Sichuan province. Several populations located in between, such as populations MYL, BX, MEK and EBY showed genetic admixture between the two clusters, indicating regions in between may serve as a contact zone. As expected, the Sichuan Basin is currently a geographic and genetic barrier for *B. albosinensis*, consistent with some previous studies (Qiu, et al. 2009, Wei, et al. 2016). Interestingly, the ENM showed that the Sichuan Basin has been unsuitable for *B. albosinensis* since the LGM, suggesting that it may be partially responsible for genetic differentiation of *B. albosinensis* populations located on both sides of the Sichuan Basin. The absence of

417 *Betula* pollen records from the Sichuan Basin during LGM further indicates the
418 inhospitable habitats for *Betula* species (Fig. 1). Although *Betula* pollen records
419 appeared during the Holocene in Sichuan Basin, it was possibly dispersed from
420 adjacent mountainous ranges, or may belong to *B. luminifera*, which currently has a
421 broad distribution around the Sichuan Basin. In contrast, for *B. albosinensis*, the
422 Qinling-Daba Mountains seems to be a dispersal corridor as populations at both sides
423 of the Qinling-Daba Mountains were not divided into genetically different groups (Fig.
424 2). Similar results have been observed for other plant species, such as *Fagus sylvatica*
425 (Magri, et al. 2006) and *Fraxinus mandshurica* (Hu, et al. 2008). If mountainous areas
426 are barriers to dispersal, it may depend on several aspects: orientation of mountain
427 range and species dispersal ability (Reeves and Richards 2014). The Taihang
428 Mountains and Lvliang Mountains in north China are of a north-south orientation and
429 the Oinling-Daba Mountains are of a west-east orientation, facilitating the dispersal of
430 *B. albosinensis* along the mountains. Furthermore, the ability of *B. albosinensis* to
431 disperse across the mountains is confirmed by our own observations that it grows on
432 the top of mountains in north China. In general, mountain ranges are more likely to be
433 dispersal barriers for lowland plant species with low dispersal ability. In our case of *B.*
434 *albosinensis*, its occupation of high altitude and strong dispersal via pollen and seed
435 could allow this species to easily spread across mountains at a broad scale. This
436 explains the low genetic differentiation among populations from north and central
437 China. The northern cluster and the southern cluster of *B. albosinensis* may also
438 reflect local adaptation to different environments, as has been reported for *Quercus*

aquifolioides (Du, et al. 2020). It is noteworthy that every individual from the southern cluster has genetic traces from the northern cluster and vice versa, a pattern mirroring incomplete lineage sorting.

Geographic pattern of genetic diversity

A key finding of our study is that the spatial pattern of genetic diversity of *B. albosinensis* can best be explained by distance from the southern edge and distance from the range edge. Our results partially support the CMH and largely fit the proposed pattern of genetic diversity(Guo 2012). However, the latitudinal effect on genetic diversity is opposite that from most previous studies(Guo 2012). In our study, genetic diversity shows a latitudinal increase whereas many other studies show a latitudinal decrease in genetic diversity. For example, a latitudinal decrease of genetic diversity has been reported for tree species distributed in NEC and subtropical areas due to northward colonisation events (Ye, et al. 2019). Our study also differs with the proposed pattern that genetic diversity decreases from the geographic centre towards both the southern and the northern range (Wei, et al. 2016). Two mutually inclusive hypotheses likely explain a latitudinal increase in genetic diversity: the existence of the northern LGM refugia and a southward movement of *B. albosinensis*. The existence of northern refugia is very likely for *B. albosinensis* as evidenced by ENM results and pollen records and a southward colonisation is also likely due to the dispersal corridor of the Qinling-Daba Mountains.

Interestingly, population HHG, located close to the geographic centre, harbors the highest level of genetic diversity, possibly due to a high abundance of *B. albosinensis*. Based on our field observations, population HHG seemed to be much larger than other populations and had numerous stands of *B. albosinensis*. Hence, we suggest that population abundance may partly account for genetic diversity for *B. albosinensis* in the Qinling Mountains. Although the ENM and pollen records indicate that northwestern Yunnan also served as LGM refugia, populations there had low genetic diversity. Two hypotheses may explain such a pattern: isolated small populations with low species abundance and a southward or an eastward movement of *B. albosinensis*. Unlike populations from the Qinling Mountains where *B. albosinensis* is the dominant species and formed very large and near pure stands in the HHG population, *B. albosinensis* from northwestern Yunnan grows sparsely among other tree species. This may result in the reduced genetic diversity due to genetic drift. Another possibility is that a southward or eastward movement of *B. albosinensis* from the geographic centre or southeastern Tibet, respectively, resulting in loss of genetic diversity due to bottleneck effects. Further inclusion of some populations from Tibet may help to clarify these points. The genetic diversity of *B. albosinensis* is also positively correlated with longitude, with higher levels of genetic diversity residing in more eastern populations. This may indicate a westward dispersal of *B. albosinensis* from central China along the Qinling-Daba Mountains.

Our work also shows that past climate by itself does not explain the patterns of

483 genetic diversity of *B. albosinensis*. For example, the model with past climate stability
484 and variability and model with current climate suitability and climatic distance had
485 low support, with AIC weights being zero and 0.001, respectively. This is somewhat
486 counterintuitive given that *B. albosinensis* seems a demanding species for moisture
487 and temperature. This is possibly due to the fact that the geographic centre of *B.*
488 *albosinensis* coincides with its climatic centre whereas its peak climatic suitability is
489 not in the geographic centre of the range but in other regions, in particular the western
490 Sichuan and northwestern Yunnan province. Another plausible explanation is that
491 distribution of suitable habitats for *B. albosinensis* did not vary much since the LGM
492 as evidenced by our ENM results. In conclusion, genetic diversity could be related to
493 abundance whereas suitability from ENMs tends not to be associated with patterns of
494 abundance (Dallas and Hastings 2018), resulting in a mismatch between genetic
495 diversity and climatic suitability.

496

497 **Acknowledgement**

498 This work was funded by the National Natural Science Foundation of China
499 (31600295 and 31770230) and Funds of Shandong 'Double Tops' Program
500 (SYL2017XTTD13).

501

502 **Author contributions**

503 NW and AD conceived the project. LL, LW, FW and NW collected samples. LL
504 carried out lab work. LL and AD analyzed the results. NW and AD edited the draft.

505

Reference

Anamthawat-Jónsson, K. and Thórsson, A. T. 2003. Natural hybridisation in birch: triploid hybrids between *Betula nana* and *B. pubescens*. - Plant Cell, Tissue Organ Cult. 75: 99-107.

Ashburner, K. and McAllister, H. A. 2016. The genus *Betula*: a taxonomic revision of birches. - Kew Publishing.

Bivand, R. and Rundel, C. 2018. rgeos: interface to geometry engine-open source (GEOS). R package version 0.3-28.

Bona, A. et al. 2018. Unfavourable habitat conditions can facilitate hybridisation between the endangered *Betula humilis* and its widespread relatives *B. pendula* and *B. pubescens*. - Plant Ecol. Divers. 11: 295-306.

Bruvo, R. et al. 2004. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. - Mol. Ecol. 13: 2101-2106.

Burgman, M. A. and Fox, J. C. 2003. Bias in species range estimates from minimum convex polygons: implications for conservation and options for improved planning. - Anim. Conserv. 6: 19-28.

Cao, X. Y. et al. 2014. Spatial and temporal distributions of major tree taxa in eastern continental Asia during the last 22,000 years. - Holocene 25: 79-91.

Chen, K. et al. 2008. Phylogeography of *Pinus tabulaeformis* Carr. (Pinaceae), a dominant species of coniferous forest in northern China. - Mol. Ecol. 17: 4276-4288.

Clark, L. V. and Jasieniuk, M. 2011. POLYSAT: an R package for polyploid microsatellite analysis. - Mol. Ecol. Resour. 11: 562-566.

D'Orazio, M. 2015. Integration and imputation of survey data in R: the StatMatch package. - Int. Stat. Rev. 63: 57-68.

Dallas, T. et al. 2017. Species are not most abundant in the centre of their geographic range or climatic niche. - Ecol. Lett. 20: 1526-1533.

Dallas, T. A. and Hastings, A. 2018. Habitat suitability estimated by niche models is largely unrelated to species abundance. - Glob. Ecol. Biogeogr. 27: 1448-1456.

Du, F. K. et al. 2020. Contrasted patterns of local adaptation to climate change across

the range of an evergreen oak, *Quercus aquifolioides*. - Evol. Appl. 13: 2377-2391.

Earl, D. A. and vonHoldt, B. M. 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. - Conserv. Genet. Resour. 4: 359-361.

Eckert, C. G. et al. 2008. Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. - Mol. Ecol. 17: 1170-1188.

Eidesen, P. B. et al. 2015. Comparative analyses of plastid and AFLP data suggest different colonization history and asymmetric hybridization between *Betula pubescens* and *B. nana*. - Mol. Ecol. 24: 3993-4009.

El Mousadik, A. and Petit, R. J. 1996. High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. - Theor. Appl. Genet. 92: 832-839.

Evanno, G. et al. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. - Mol. Ecol. 14: 2611-2620.

Fick, S. E. and Hijmans, R. J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. - Int. J. Climatol. 37: 4302-4315.

Gent, P. R. et al. 2011. The Community Climate System Model version 4. - J. Clim. 24: 4973-4991.

Goudet, J. 1995. FSTAT (version 1.2): a computer program to calculate F-statistics. - Heredity (Edinb) 86: 485-486.

Gougherty, A. V. et al. 2020. Contemporary range position predicts the range-wide pattern of genetic diversity in balsam poplar (*Populus balsamifera* L.). - J. Biogeogr. 47: 1246-1257.

Guo, Q. 2012. Incorporating latitudinal and central–marginal trends in assessing genetic variation across species ranges. - Mol. Ecol. 21: 5396-5403.

Guo, Y. et al. 2019. Canopy disturbance and gap partitioning promote the persistence of a pioneer tree population in a near-climax temperate forest of the Qinling Mountains, China. - Ecol. Evol. 9: 7676-7687.

Hao, Q. et al. 2018. The critical role of local refugia in postglacial colonization of

564 Chinese pine: joint inferences from DNA analyses, pollen records, and species
565 distribution modeling. - Ecography 41: 592-606.
566 Harrison, S. P. et al. 2001. The role of dust in climate changes today, at the last glacial
567 maximum and in the future. - Earth Sci. Rev. 54: 43-80.
568 Hewitt, G. M. 1999. Post-glacial re-colonization of European biota. - Biol. J. Linn.
569 Soc. Lond. 68: 87-112.
570 Hewitt, G. M. 2004. Genetic consequences of climatic oscillations in the Quaternary. -
571 Philos. Trans. R. Soc. Lond., B, Biol. Sci. 359: 183-195.
572 Hijmans, R. J. et al. 2005. Very high resolution interpolated climate surfaces for
573 global land areas. - Int. J. Climatol. 25: 1965-1978.
574 Hou, Z. et al. 2018. Phylogeographic analyses of a widely distributed *Populus*
575 *davidiana*: further evidence for the existence of glacial refugia of cool-temperate
576 deciduous trees in northern east Asia. - Ecol. Evol. 8: 13014-13026.
577 Hu, L. J. et al. 2008. Nuclear DNA microsatellites reveal genetic variation but a lack
578 of phylogeographical structure in an endangered species, *Fraxinus mandshurica*,
579 across north-east China. - Ann. Bot. 102: 195-205.
580 Hu, Y. N. et al. 2019. Population structure of *Betula albosinensis* and *Betula*
581 *platyphylla*: evidence for hybridization and a cryptic lineage. - Ann. Bot. 123:
582 1179-1189.
583 Jin, P. Y. et al. 2020. Geography alone cannot explain *Tetranychus truncatus* (Acari:
584 Tetranychidae) population abundance and genetic diversity in the context of the
585 center-periphery hypothesis. - Heredity (Edinb) 124: 383-396.
586 Kindt, R. 2018. Ensemble species distribution modelling with transformed suitability
587 values. - Environ. Model. Softw. 100: 136-145.
588 Kulju, K. K. M. et al. 2004. Twenty-three microsatellite primer pairs for *Betula*
589 *pendula* (Betulaceae). - Mol. Ecol. Notes 4: 471-473.
590 Lichstein, J. W. et al. 2002. Spatial autocorrelation and autoregressive models in
591 ecology. - Ecol. Monogr. 72: 445-463.
592 Lira-Noriega, A. and Manthey, J. D. 2014. Relationship of genetic diversity and niche
593 centrality: a survey and analysis. - Evol. 68: 1082-1093.

Liu, J. Q. et al. 2012. Phylogeographic studies of plants in China: advances in the past and directions in the future. - J. Syst. Evol. 50: 267-275.

Magri, D. et al. 2006. A new scenario for the quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. - New Phytol. 171: 199-221.

Micheletti, S. J. and Storfer, A. 2015. A test of the central-marginal hypothesis using population genetics and ecological niche modelling in an endemic salamander (*Ambystoma barbouri*). - Mol. Ecol. 24: 967-979.

Nei, M. 1987. Molecular evolutionary genetics. - Columbia university press.

Ortego, J. et al. 2015. Climatically stable landscapes predict patterns of genetic structure and admixture in the Californian canyon live oak. - J. Biogeogr. 42: 328-338.

Petit, R. et al. 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. - Science 300: 1563-1565.

Petit, R. J. et al. 1997. Chloroplast DNA footprints of postglacial recolonization by oaks. - PNAS 94: 9996-10001.

Pritchard, J. K. et al. 2000. Inference of population structure using multilocus genotype data. - Genetics 155: 945-959.

Qian, H. and Ricklefs, R. E. 2001. Palaeovegetation (Communications arising): diversity of temperate plants in east Asia. - Nature 413: 130-131.

Qiu, Y. X. et al. 2009. Did glacials and/or interglacials promote allopatric incipient speciation in east Asian temperate plants? Phylogeographic and coalescent analyses on refugial isolation and divergence in *Dysosma versipellis*. - Mol. Phylogenet. Evol. 51: 281-293.

Qiu, Y. X. et al. 2011. Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. - Mol. Phylogenet. Evol. 59: 225-244.

R Core Team 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reeves, P. A. and Richards, C. M. 2014. Effect of a geographic barrier on adaptation

624 in the dwarf sunflower (*Helianthus pumilus* Nutt.). - Int. J. Plant Sci. 175: 688-701.
625 Rodríguez, C. A. and Pateiro, L. B. 2010. Generalizing the convex hull of a sample:
626 the R package alphahull. - J. Stat. Softw.
627 Sagarin, R. D. and Gaines, S. D. 2002. The 'abundant centre' distribution: to what
628 extent is it a biogeographical rule? - Ecol. Lett. 5: 137-147.
629 Schielzeth, H. 2010. Simple means to improve the interpretability of regression
630 coefficients. - Methods Ecol. Evol. 1: 103-113.
631 Swaegers, J. et al. 2013. Rapid range expansion increases genetic differentiation while
632 causing limited reduction in genetic diversity in a damselfly. - Heredity (Edinb) 111:
633 422-429.
634 Tian, B. et al. 2009. Phylogeographic analyses suggest that a deciduous species
635 (*Ostryopsis davidiana* Decne., Betulaceae) survived in northern China during the Last
636 Glacial Maximum. - J. Biogeogr. 36: 2148-2155.
637 Truong, C. et al. 2005. Isolation and characterization of microsatellite markers in the
638 tetraploid birch, *Betula pubescens* ssp. *tortuosa*. - Mol. Ecol. Notes 5: 96-98.
639 Tsuda, Y. et al. 2008. Development of 14 EST-SSRs for *Betula maximowicziana* and
640 their applicability to related species. - Conserv. Genet. 10: 661-664.
641 Tsuda, Y. et al. 2017. Multispecies genetic structure and hybridization in the *Betula*
642 genus across Eurasia. - Mol. Ecol. 26: 589-605.
643 Wagenmakers, E. J. and Farrell, S. 2004. AIC model selection using Akaike weights. -
644 Psychon. Bull. Rev. 11: 192-196.
645 Wang, N. et al. 2013. Genome sequence of dwarf birch (*Betula nana*) and
646 cross-species RAD markers. - Mol. Ecol. 22: 3098-3111.
647 Wang, N. et al. 2014. Molecular footprints of the Holocene retreat of dwarf birch in
648 Britain. - Mol. Ecol. 23: 2771-2782.
649 Wang, N. et al. 2016. Molecular phylogeny and genome size evolution of the genus
650 *Betula* (Betulaceae). - Ann. Bot. 117: 1023-1035.
651 Wang, W. T. et al. 2016. Phylogeography of postglacial range expansion in *Juglans*
652 *mandshurica* (Juglandaceae) reveals no evidence of bottleneck, loss of genetic
653 diversity, or isolation by distance in the leading-edge populations. - Mol. Phylogenet.

654 Evol. 102: 255-264.

655 Warren, D. L. et al. 2010. ENMTools: a toolbox for comparative studies of
656 environmental niche models. - Ecography 33: 607-611.

657 Wei, X. Z. et al. 2016. Genetic evidence for central-marginal hypothesis in a Cenozoic
658 relict tree species across its distribution in China. - J. Biogeogr. 43: 2173-2185.

659 Wu, B. et al. 2002. Development of microsatellite markers in white birch (*Betula*
660 *platyphylla* var. *japonica*). - Mol. Ecol. Notes 2: 413-415.

661 Yannic, G. et al. 2013. Genetic diversity in caribou linked to past and future climate
662 change. - Nat. Clim. Chang. 4: 132-137.

663 Ye, J. W. et al. 2019. Phylogeography of *Schisandra chinensis* (Magnoliaceae) reveal
664 multiple refugia with ample gene flow in northeast China. - Front. Plant Sci. 10: 199.

665 Zeng, Y. F. et al. 2015. Multiple glacial refugia for cool-temperate deciduous trees in
666 northern east Asia: the Mongolian oak as a case study. - Mol. Ecol. 24: 5676-5691.

667 Zhang, Q. et al. 2005. Phylogeography of the Qinghai-Tibetan Plateau endemic
668 *Juniperus przewalskii* (Cupressaceae) inferred from chloroplast DNA sequence
669 variation. - Mol. Ecol. 14: 3513-3524.

670

671 **Conflicts of interest**

672 There is no conflict of interest in this study.

673

Figure legends

Figure 1 The predicted climatic suitability for *B. albosinensis* and pollen records of *Betula* species. (a) Suitable climate during the LGM, (b) the mid-Holocene, (c) present and (d) in the future (2070). Orange and blue points represent pollen records of *Betula* species during LGM and the Holocene, respectively. Red points represent coordinate points used for ecological niche modelling.

Figure 2 STRUCTURE results for K=2 based on microsatellite markers and a map showing sampling sites. Pie charts indicate the admixture of each population.

Figure 3 A principal component analysis (PCoA) of *B. albosinensis* at 16 microsatellite markers. Green, yellow and blue points represent "northern individuals", "southern individuals" and admixed individuals between the northern cluster and the southern cluster, respectively.

Figure 4 Maps and correlograms (a, b) of genetic diversity (expected heterozygosity = *He*; gene diversity = *Gd*; allele richness = *Ar*) among *B. albosinensis* populations. Circle size in the correlograms is proportional to the number of records within each distance class and filled circles indicate significant autocorrelation at particular distance class (two sided, $p > .975$ or $p < .025$).

Figure 5 Linear model fit of genetic diversity (expected heterozygosity = *He*; gene diversity = *Gd*; allele richness = *Ar*) with latitude and longitude, respectively.

Figure S1 The optimal number of clusters inferred using the "Evanno test" method (A) and (B) STRUCTURE results at K values between 2 and 6 based on microsatellite markers.

696 **Table legends**

697 **Table 1** Summary statistics for conditional autoregressive models for genetic diversity

698 (*He*, *Gd* and *Ar*) in *B. albosinensis*.

699 **Table S1** Detailed information of populations used in the present study.

700 **Table S2** Detailed information of pollen records used in the present study.

701

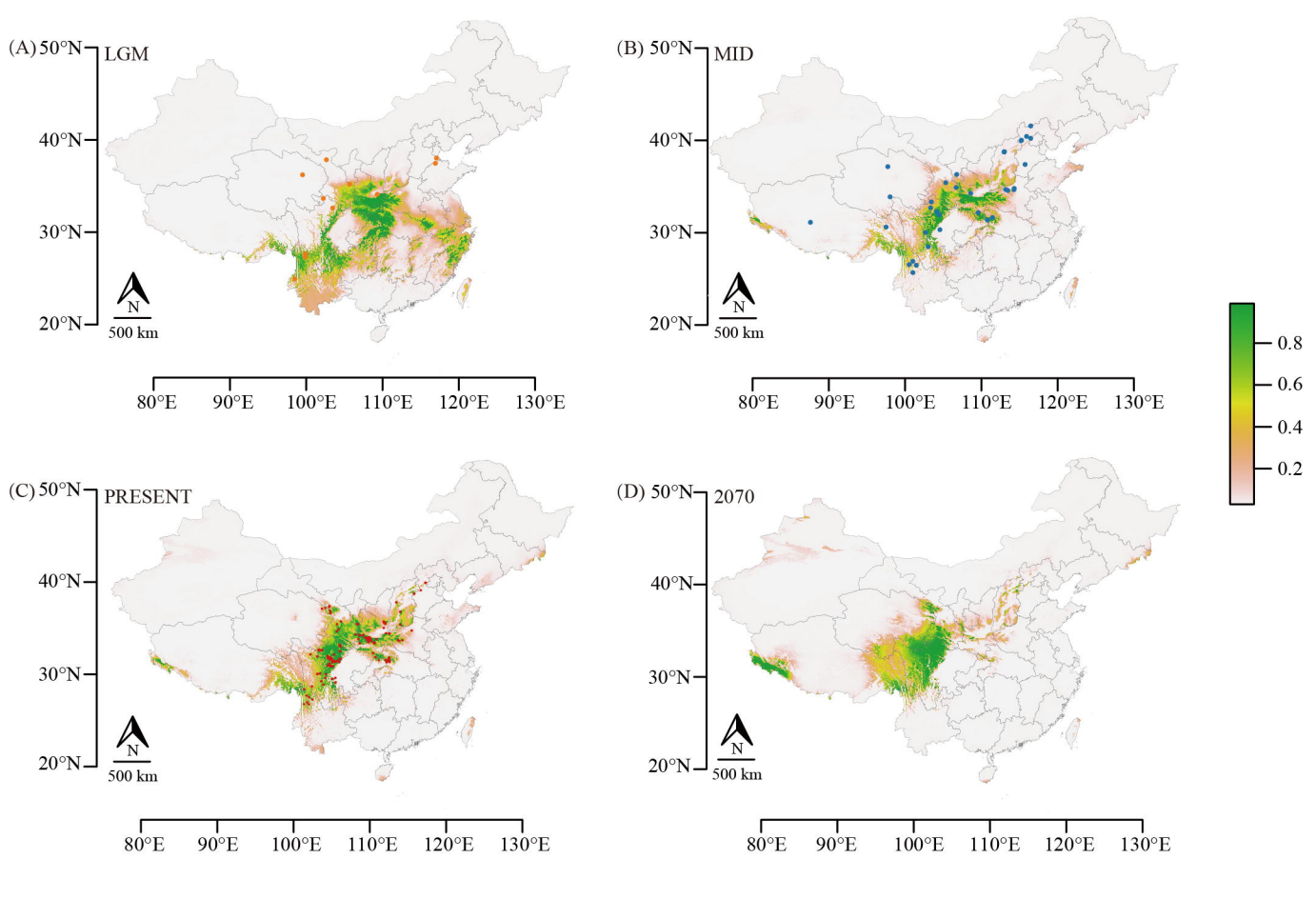
702

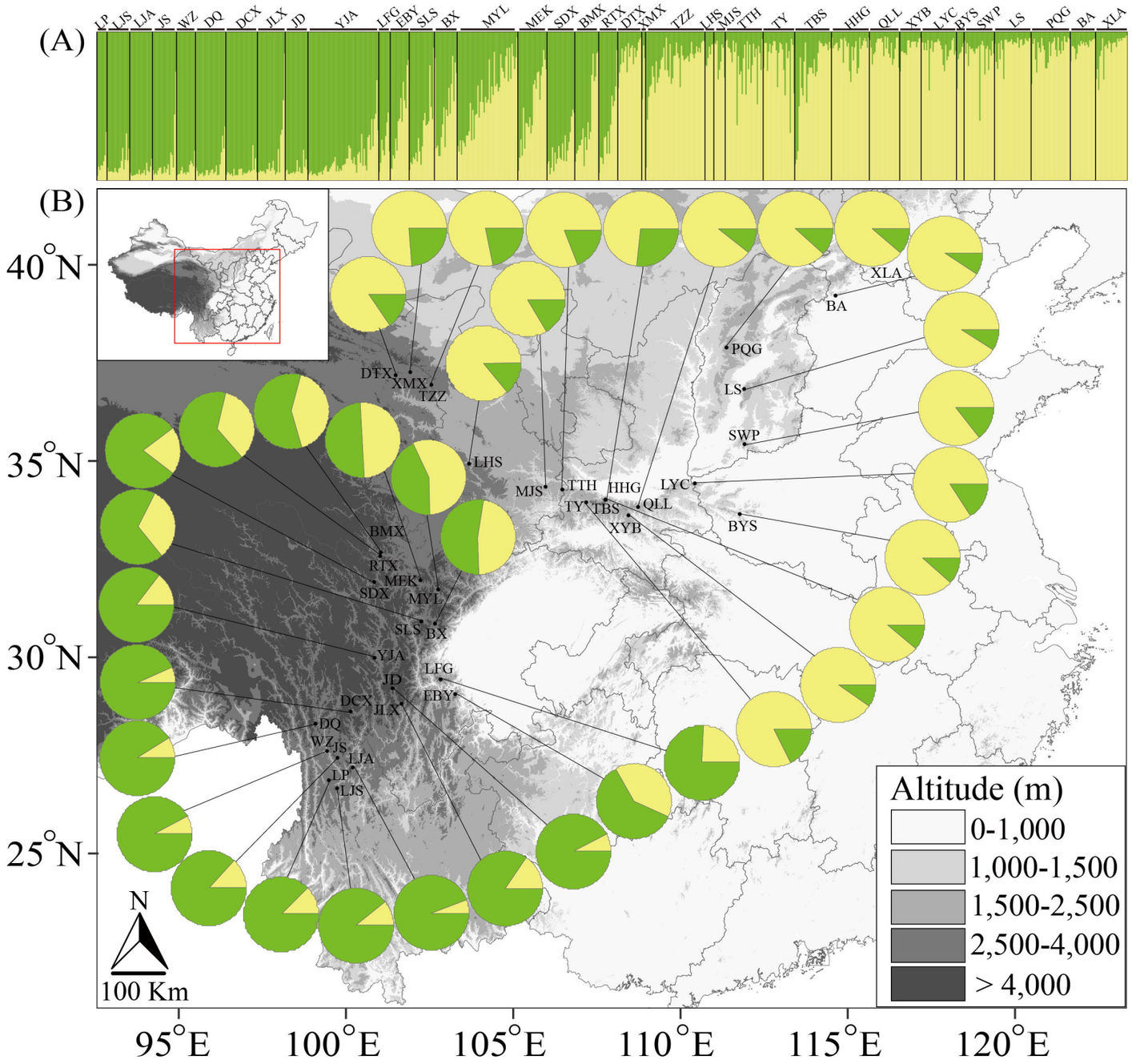
703 **Table 1**. Summary statistics for conditional autoregressive models for a range-wide
704 sample of *He*, *Gd* and *Ar* in *Betula albosinensis*, ranked by relative support.

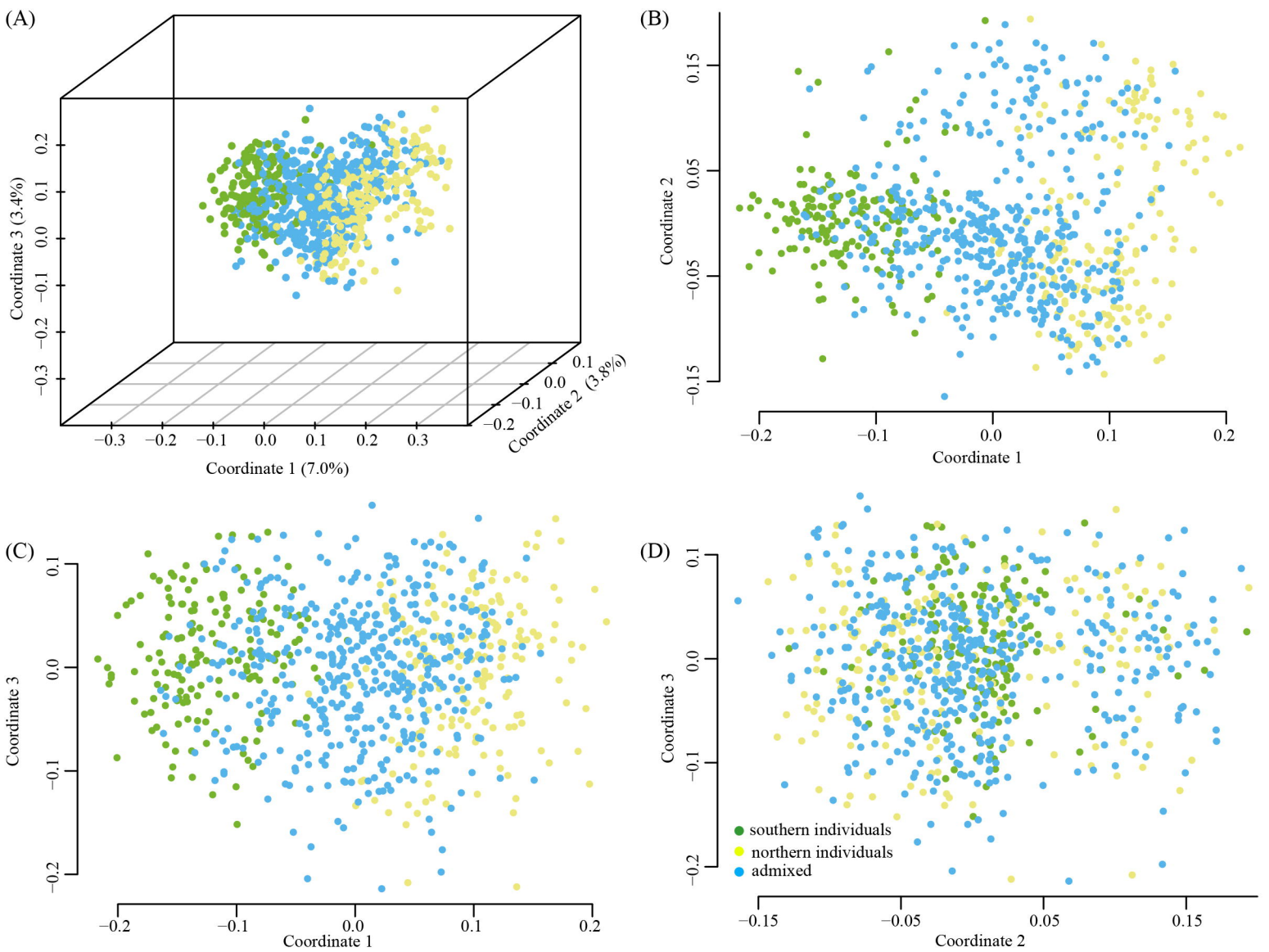
	He			*Gd*		
Model	Nagelkerke R^2	AIC	AIC weight	Nagelkerke R^2	AIC	AIC weight
southernEdge+GeoEdge	0.464	-171.099	0.978	0.644	-200.621	0.997
southernEdge	0.289	-162.655	0.014	0.479	-188.459	0.002
geoEdge+geoCenter	0.291	-160.769	0.006	0.417	-182.317	0.000
suitability+climDist	0.184	-157.946	0.001	0.253	-182.793	0.000
admixture+maxCluster	0.227	-155.560	0.000	0.443	-182.008	0.000
stability+variability	0.067	-155.361	0.000	0.070	-178.550	0.000

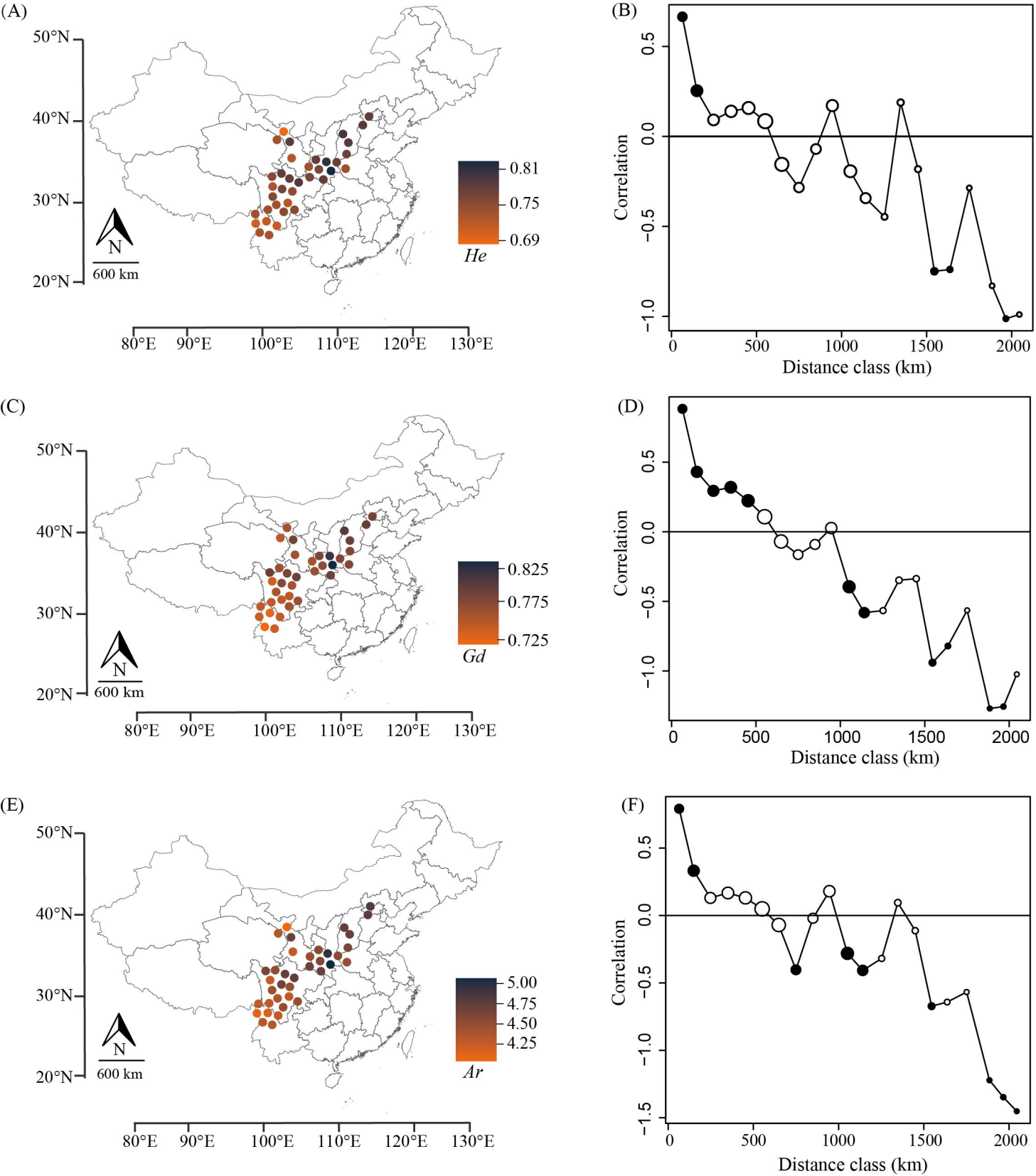
705

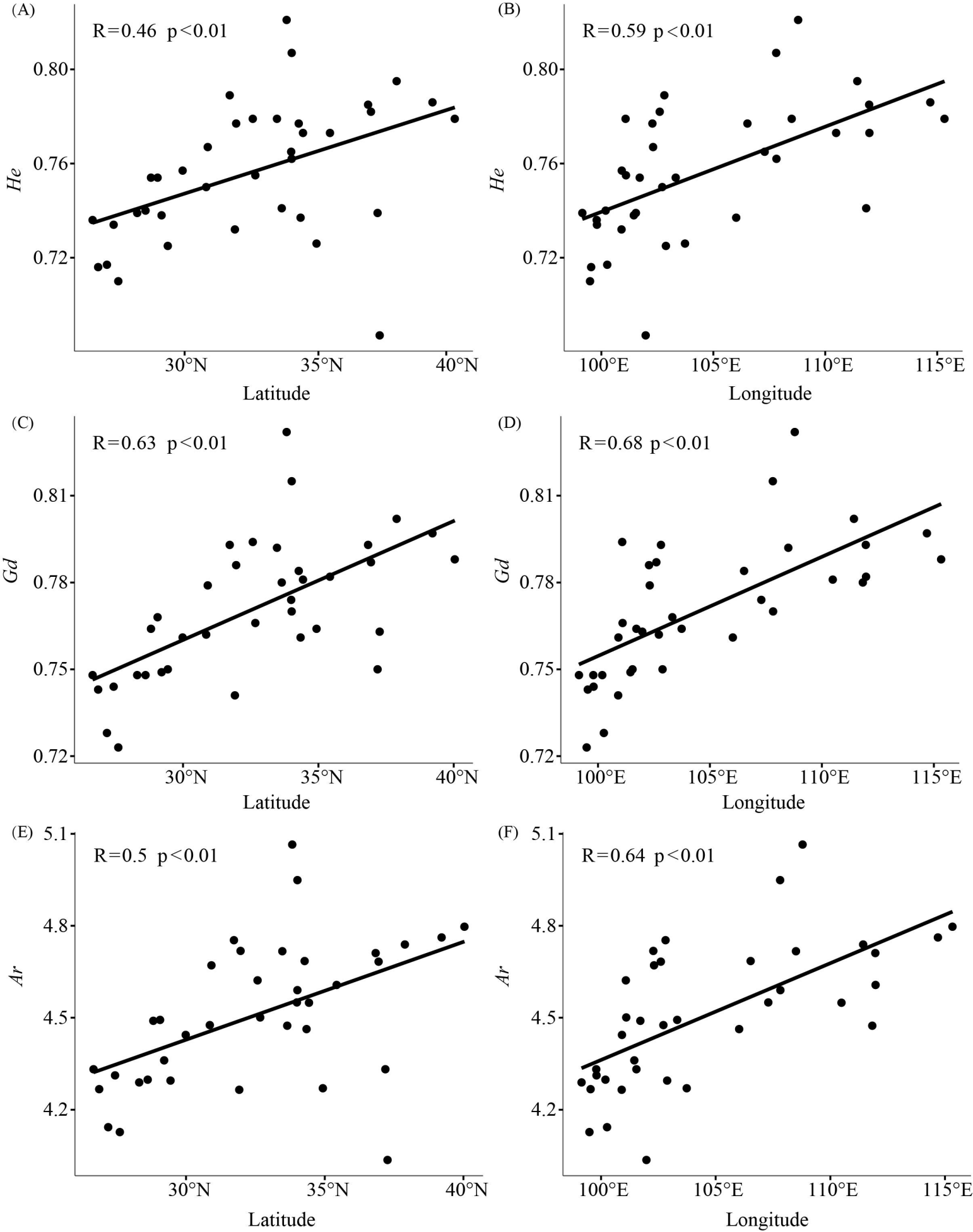
706


(A) LGM
(B) MID
(C) PRESENT
(D) 2070
50°N
40°N
30°N
20°N
80°E
90°E
100°E
110°E
120°E
130°E
N
500 km
0.8
0.6
0.4
0.2




(A)
Coordinate 3 (3.4%)
Coordinate 1 (7.0%)
Coordinate 2 (3.8%)
(B)
Coordinate 2
Coordinate 1
(C)
Coordinate 3
Coordinate 1
(D)
Coordinate 3
Coordinate 2
southern individuals
northern individuals
admixed


(A)
50°N
40°N
30°N
20°N
N
600 km
0.81
0.75
0.69
He
80°E
90°E
100°E
110°E
120°E
130°E
(B)
Correlation
0.5
0.0
−0.5
−1.0
0
500
1000
1500
2000
Distance class (km)
(C)
50°N
40°N
30°N
20°N
N
600 km
0.825
0.775
0.725
Gd
80°E
90°E
100°E
110°E
120°E
130°E
(D)
Correlation
0.5
0.0
−0.5
−1.0
0
500
1000
1500
2000
Distance class (km)
(E)
50°N
40°N
30°N
20°N
N
600 km
5.00
4.75
4.50
4.25
Ar
80°E
90°E
100°E
110°E
120°E
130°E
(F)
Correlation
0.5
0.0
−0.5
−1.0
−1.5
0
500
1000
1500
2000
Distance class (km)


(A)
R=0.46 p<0.01
He
0.80
0.76
0.72
30°N
35°N
40°N
Latitude
(B)
R=0.59 p<0.01
He
0.80
0.76
0.72
100°E
105°E
110°E
115°E
Longitude
(C)
R=0.63 p<0.01
Gd
0.81
0.78
0.75
0.72
30°N
35°N
40°N
Latitude
(D)
R=0.68 p<0.01
Gd
0.81
0.78
0.75
0.72
100°E
105°E
110°E
115°E
Longitude
(E)
R=0.5 p<0.01
Ar
5.1
4.8
4.5
4.2
30°N
35°N
40°N
Latitude
(F)
R=0.64 p<0.01
Ar
5.1
4.8
4.5
4.2
100°E
105°E
110°E
115°E
Longitude