Phylogeography of *Amygdalus mongolica* in relation to Quaternary climatic aridification and oscillations in northwestern China

Lin Zhang [1, *], Fangfang Sun [1, *], Songmei Ma [1, #], Chuncheng Wang [2], Bo Wei [3], Yunling Zhang [4]

[1] Shihezi University, Xinjiang Production and Construction Corps Key Laboratory of Oasis Town and Mountain-basin System Ecology, College of Science, Shihezi, Xinjiang, China.

[2] Shihezi University, Xinjiang Production and Construction Corps Key Laboratory of Oasis Town and Mountain-basin System Ecology, College of Life Sciences, Shihezi, Xinjiang, China.

[3] Institute of Geographic Sciences and Natural Resources Research, Beijing, China.

[4] General grassland station of Xinjiang, Urumqi, Xinjiang, China.

#Corresponding author: Songmei Ma, shzmsm@126.com

* Lin Zhang and Fangfang Sun contribute equally to this work.

Abstract

Quaternary period geological events and climatic oscillations significantly affect the geographic structure and genetic diversity of species distribution in arid northwestern China. *Amygdalus mongolica* is a relict and endangered shrub that occurs primarily in arid areas of northwestern China. Based on variation patterns present at three cpDNA regions (*psb*K-*psb*I, *trn*L-*trn*F and *trn*V) and in one nDNA sequence (ITS1-ITS4) in 174 individuals representing 15 populations, the spatial genetic structure and demographic history of *A. mongolica* was examined across its entire geographic range. The 17 different haplotypes and 10 ribotypes showed two lineages, distributed across the Western (Mazong Mountains, Hexi Corridor, and Alxa Left Banner) and Eastern regions (Urad Houqi, Yinshan Mountains, Urad Zhongqi, and Daqing Mountains) according to the median-joining network and the BI (Bayesian inference) and ML (Maximum likelihood) trees. AMOVA analysis demonstrated that over 65% of the observed genetic variation was related to this lineage split. The expansions of the Ulanbuhe and Tengger deserts and the eastward extension of the Yinshan Mountains since the Quaternary period likely interrupted gene flow and triggered the observed divergence in the two allopatric regions; arid landscape fragmentation accompanied by local environmental heterogeneity further increased local adaptive differentiation between the Western and Eastern groups. Based on the evidence from phylogeographical patterns and the distribution of genetic variation, *A. mongolica* distributed in the eastern and western regions are speculated to have experienced eastward migration along the southern slopes of the Lang Mountains and westward migration along the margins of the Ulanbuhe and Tengger deserts to the Hexi

31   Corridor, respectively. For setting a conservation management plan, it is recommended that the south slopes of

32   the Lang Mountains and northern Helan Mountains be identified as the two primary conservation areas, as they

33   have high genetic variation and habitats that are more suitable.

35   **Keywords** *Amygdalus mongolica*, allopatric differentiation, post-glacial expansion, genetic conservation,

36   northwestern China

37   **Introduction**

38   Climatic oscillations, especially those that occurred in the Pleistocene glacial-interglacial cycles during

39   the Quaternary period, shaped contraction and expansion patterns of species, and undoubtedly left genetic

40   signatures in extant populations (Arbogast 2001; Hewitt 1996). Many plant phylogeographic studies using the

41   combined approach of molecular data and paleoclimatic evidence have found that Pleistocene glaciations had a

42   profound influence on the genetic diversity, population structure, and evolutionary history of species

43   throughout the Northern Hemisphere (Hewitt 2000; Riccioni et al. 2019). In the arid region of northwest

44   China, although pollen records and paleodata show no evidence of glaciation (Shi 2006), Quaternary climatic

45   shocks have also profoundly affected the local plant growth, during the Last Glacial Maximum, the current

46   coniferous and deciduous forests were replaced by steppe and even desert vegetation in northern and

47   northwestern China (Harrison et al. 2001; Ni et al. 2006; Yu et al. 2000). Aridification occurred in northwest

48   China due to the uplifts of the Tibet Plateau and the difficulty of water vapor reaching inland areas, and further

49   intensified as desert conditions began to emerge during the late Miocene (Zhang & Jiang 1992). At the same

50   time, extremely arid climates and geological events created fragmented landscapes in northwestern China,

51   potentially having led to the reduction of suitable plant habitat in these areas (Jia & Zhang 2021; Li et al. 2019;

52   Qiu et al. 2017). The fragmented landscape in arid northwestern China during the Quaternary period restricted

53   gene flow between different geographical populations, and has played significant roles in determining

54   divergence and diversification of local desert plants in isolated and allopatric regions. Currently, there is

55   significant regional divergence in xerophytic *Tugarinovia mongolica* from Inner Mongolia due to habitat

56   fragmentation following enhanced aridification (Zhao et al. 2019a). Intraspecific differentiation and the deeply

57   shaped genetic structure for *Gymnocarpos przewalskii* and *Lagochilus ilicifolius* were primarily caused by

58 variation of desert habitats during the Pleistocene (Gao et al. 2014; Zhang et al. 2020). In contrast, xerophytic

59 plants such as *Zygophyllum xanthoxylon* and *Clematis songorica* appeared to adapt to the changes in aridity,

60 and no evidence of genetic divergence among different populations was detected (Shi & Zhang 2015; Zhang et

61 al. 2013). Several refugia for several local desert species have been revealed in fragmented and arid areas in

62 the Tarim Basin, Tianshan Mountains, Hami Basin, Helan Mountains and northwestern areas of Inner

63 Mongolia in northwestern China, and evidence of post-glacial expansion following glacial periods has been

64 detected (Shi & Zhang 2015; Xu & Zhang 2015; Zeng et al. 2018). The arid Tianshan Mountains have greatly

65 affected many plants genetic structure, such as *Clematis sibirica* of a woody perennial vine growing primarily

66 under conifer forests (Zhang & Zhang 2012) and *Delphinium naviculare* growing at forest edges and in grassy

67 slopes (Zhang et al. 2013); the Helan Mountains likely acted as the center of diversification for *Lagochilus*

68 *ilicifolius* (Meng & Zhang 2011) and were inferred as the refugial locations for *Zygophyllum xanthoxylon* (Shi

69 & Zhang 2015).

70   *Amygdalus mongolica* is a xerophitic desert shrub that ranges from the Mazong Mountains in the

71 northernmost Gansu, through the Hexi Corridor and Alxa Left Banner to the Urad Houqi, Yinshan Mountains,

72 Urad Zhongqi, and Daqing Mountains in northeastern Inner Mongolia in northwestern China (Fu & Chin

73 1992). The area for this study was concentrated in areas of severe aridity, characterized by low annual

74 precipitation and high evaporation (Ma et al. 2015), where altitude was greater than 1000m and the terrain was

75 mainly plateaus (Alxa Plateau and Ordos Plateau) and desert landscapes (Badain Jaran Desert, Tengger Desert,

76 Ulan Buhe Desert and Kubuqi desert). The Ordos Plateau is surrounded by the Yin Mountains in an east-west

77 direction with many deep valleys, and the Helan Mountains in a south-north direction that has many valleys

78 and steep terrain. *A. mongolica*, as a Tertiary Miocene relic plant of the Ancient Mediterranean, offers an

79 optimal case for investigating the evolutionary processes in response to the Quaternary climate oscillations and

80 aridification in arid northwestern China (Fu & Chin 1992). Moreover, it is endangered (listed on the China

81 Species Red List) and is a protected plant of third conservation priority (Fu & Chin 1992). Because of this, the

82 genetic structure of the *A. mongolica* population has high importance, for understanding both the origin and

83 evolution of plants in the Chinese northwestern desert, as well as for plant protection and conservation. In a

84 previous study by our group, , the presence of two chloroplast intergenic spacers (*psb*K-*psb*I and *trn*L-*trn*F)

85 supported two fragmented geographic groups among different populations of *A. mongolica* (Ma et al. 2019). A

86 large number of shared haplotypes exist between the two geographic groups, however, likely because of

87 incomplete lineage sorting; the limited polymorphic fragments were deficient for studying population

88 dynamics, genetic structure, and conservation of species genetic diversity.

89   Previous studies have shown that a variety of genes, such as maternally inherited chloroplast DNA

90 (cpDNA) markers in combination with biparentally inherited nuclear DNA (nDNA) markers, show an integral

91 view for identifying the relationships and genetic structure among taxonomic groups and populations (Li et al.

92 2019). Here, multiple methods of population genetic analysis and landscape genetic analysis were used along

93 with a least-cost path (LCP) analysis to investigate the genetic structure and demographic history of 15 natural

94 populations of *A. mongolica* in northwestern China based on three highly polymorphic chloroplast DNA

95 fragments (*psb*K-*psb*I, *trn*L-*trn*F and *trn*V) and a nuclear marker (ITS1-ITS4). This study aimed to assess the

96 spatial genetic structure and intraspecific differentiation and detect whether incomplete lineage sorting exists in

97 *A. mongolica* populations in northwestern China according to cpDNA and nDNA sequences variation,

98 inference of phylogeographic and demographic history of *A. mongolica* populations during Quaternary climate

99 fluctuations and desert expansion was also performed, along with comprehensive identification of potential

100 conservation areas for *A. mongolica* populations.

101 **Material and Methods**

102 **Sample collection**

103   A total of 174 individuals of *A. mongolica* from 15 natural populations were investigated (Table 1)

104 throughout its geographic distribution in northwestern China (Fig. 1a). The details of sampling populations are

105 described in our previous study (Ma et al. 2019), including three from Gansu Province and twelve from Inner

106 Mongolia (Table 1). Fresh young leaves were collected from 7-15 individuals in each population. Individual

107 plants were sampled at least 10 m apart to avoid sampling close relatives and to ensure a robust sampling. The

108 leaves were dried on silica gel and stored at 4℃ until DNA extraction. The Provincial administrative boundary

109 data of the study area were obtained from the Data Center of Resources and Environment Sciences, Chinese

110 Academy of Sciences (http://www.resdc.cn/).

111 **Laboratory Procedures**

112   Total genomic DNA was extracted from approximately 100 mg of silica gel-dried leaf tissue using a

113   modified 2×CTAB protocol (Doyle 1987). Three polymorphic chloroplast DNA intergenic spacers, *psb*K-*psb*I,

114   *trn*L-*trn*F and *trn*V (Shaw et al. 2007) and one nDNA fragment, ITS1-ITS4 (White et al. 1990) were chosen

115   for this study.

116   We used a 30 μL volume containing 20 mM dNTP, 2 mM MgCl$_2$, 0.025 U mL^{-1} *Taq* polymerase and 1

117   pmol of each primer (Takara Co. Ltd., Beijing, China) to conduct the polymerase chain reaction (PCR) (Takara

118   Co. Ltd., Beijing, China). The DNA amplification profile was 94℃ for 2 min, followed by 30 cycles of 94℃

119   for 30 s, annealing at 52 ℃ (*psb*K-*psb*I, *trn*L-*trn*F), 54 ℃ (*trn*V) and 55 ℃ (ITS1-ITS4) for 30 s, 72 ℃ for 90 s,

120   and a final extension at 72 ℃ for 10 min. The processing method of all PCR products were previously

121   described in Ma (Ma et al. 2015). Specifically, the study used a PCR product purification kit (Viogene,

122   Sunnyvale, CA, USA) to purify from agarose gel (0.1-0.5%) and used the ABI 3730 DNA Analyzer to

123   sequence on forward and reverse strands by BigDye terminator chemistry (Applied Biosystems, Foster City,

124   CA, USA). For the heterozygous sites for nuclear genes after sequencing, using the Bayesian approach of

125   PHASE 2.1 to infer haplotypes from genotype data (Stephen & Donnelly 2003), and our study was corrected

126   using default parameters in DnaSP 6.0 and verified in MEGA-X ver. 10.1.8. Sequences were aligned using

127   CLUSTAL X1.83 (Thompson et al. 1997) and then manually used BioEdit 7.2.5 (Hall 1999) for alignment.

128   Each insertion / deletion in this study was treated as a single mutation event and encoded as substitutions in

129   subsequent analyses (Simmons & Ochoterena 2000).

130   **Genetic diversity and genetic structure analyses**

131   Molecular parameters, containing cpDNA haplotype / nDNA ribotype diversity (H_d, R_d), and nucleotide

132   diversity (π) based on cpDNA and nDNA dataset, were calculated in DNASP 6.0 (Librado & Rozas 2009).

133   The two indexes were also mapped using 'ggplot2' package in R 4.0.5

134   (https://cran.rproject.org/web/packages/ggplot2/index.html) (Wickham 2016). Total genetic diversity across all

135   populations (H_T), average genetic diversity within populations (H_S), and two parameters of population

136   differentiation (G_{ST}, N_{ST}) were calculated using HAPLONST 3.0 (Pons & Petit 1996). PERMUT 2.0 was used

137   to test the significance of population differentiation (G_{ST}, N_{ST}) (Bandelt et al. 1999; Pons & Petit 1996) with

138   1,000 replicates. Phylogeographical structure is present if N_{ST} is significantly higher than G_{ST} (Pons & Petit

139   1996).

140   An AMOVA analysis was employed to investigate genetic differentiation among overall populations and

141   the defined population groups by phylogenetic clustering results using ARLEQUIN 3.5.2.2 (Excoffier &

142   Lischer 2010). The relationships among haplotypes were estimated using the median-joining method,

143   implemented in Network 5.01 (Bandelt et al. 1999). Based on the genetic distances among populations, we

144   used Principal coordinate analyses (PCoAs) to analyze the genetic structure for *A. mongolica* using the vegan

145   package in R 4.0.5 (Oksanen et al. 2007), and then we performed 3D visualization using the scatter plot 3d

146   package in R 4.0.5. Alleles In Space (Miller 2005) was used to explore Genetic landscape shape analysis of *A.*

147   *mongolica* which can reveal the level of genetic differentiation among different populations from the

148   perspective of Landscape genetics, and the result of genetic landscape shape analysis plot which *x*- and *y*-axes

149   correspond to geographical locations and the *z*-axis to genetic distance, and the peak height on the z-axis

150   represents the degree of genetic differentiation between populations. Furthermore, Based on Monmonier's

151   maximum difference algorithm, we estimated the possible genetic barriers in BARRIER v2.2 (Manni et al.

152   2004). We implemented a multiple matrices test based on 100 replicates of population average pairwise

153   difference matrices in order to assess the robustness computed barriers. These matrices were generated by

154   resampling of genotype sequences in DnaSP 6.0 (Librado & Rozas 2009) and subsequent analyses in

155   ARLEQUIN 3.5.2.2 (Excoffier & Lischer 2010). In order to identify whether the genetic barriers and

156   biogeographic barriers were spatially consistent, we corrected the possible spatial position of genetic barriers

157   in ArcGIS 10.5 (ESRI, Redlands, CA, USA) according to the sampled population points.

158   **Phylogenetic analysis and estimation of divergence**

159   Phylogenetic trees with cpDNA haplotypes and nDNA ribotypes in *A. mongolica* were reconstructed

160   using maximum likelihood (ML) analysis with MEGA-X ver. 10.1.8 (Sudhir et al. 2018) and a Bayesian

161   inference (BI), as implemented in BEAST 2.2.1 (Drummond & Rambaut 2007) . *Amygdalus davidiana* and

162   *Prunus triloba* were selected as the outgroups in the analysis owing to their relatively close evolutionary

163   relationship (Delplancke et al. 2016). FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) was used for

164   tree-visualization (Zhao et al. 2019b).

165   The HKY substitution model was selected by MODELTEST as the best-fitting model for the data set

166   (Posada & Crandall 1998). Based on the synonymous substitution rates for most angiosperm species of

167   cpDNA genes (1.0×10^{-9} to 3.0×10^{-9} s/s/y) (Wolfe et al. 1987), and substitution rates of plant ITS ($3.46 \times$

168   10^{-9} to 8.69×10^{-9} s/s/y) (Richardson et al. 2001), we used a normal prior distribution and set a mean of $2 \times$

169   10^{-9} s/s/y and a standard deviation of 6.08×10^{-10} s/s/y for cpDNA, and a mean of 6.73×10^{-9} s/s/y and a

170   standard deviation of 1.99×10^{-9} s/s/y for ITS (Zhang et al. 2013). The Markov chain Monte Carlo (MCMC)

171   analysis was run for 10 million generations and sampled every 1000 generations for the Bayesian analysis.

172   Bayes factor (BF) values were used to detect MCMC convergence, and the effective sample size (ESS) of each

173   parameter above 200 after the first 10% of generations was discarded as burn-in (Baele et al. 2012).

174   **Demographic history analysis and potential migration corridors**

175   The mismatch distribution analysis and the parameter calculations were all implemented in ARLEQUIN

176   3.5.2.2 (Rogers & Harpending 1992). Tajima's D (Tajima 1989) and Fu's Fs statistics were calculated to test

177   the recent demographic expansion. Significantly large negative values for D and F_S suggested that populations

178   have experienced range expansion or natural selection (Fu 1997; Tajima 1989). And the Harpending's

179   raggedness index (H_{rag}), the sum of squared deviation (SSD) and their p values were computed to test the

180   significance of this population expansion model. A non-significant value (p>0.05) of SSD and H_{rag} indicated

181   population expansion (Excoffier et al. 2005; Harpending 1994). The mismatch distribution analysis was used

182   to test whether *A. mongolica* underwent a recent range expansion for the overall populations and the defined

183   groups (Excoffier & Lischer 2010). Unimodal distributions of pairwise differences suggested populations that

184   could have undergone range expansions, whereas populations in demographic equilibrium were characterized

185   by multimodal mismatch distributions (Rogers & Harpending 1992). Expansion tests with 10,000 permutations

186   were performed for the significance test. To further detect the expansion trend and estimate the expansion time

187   of *A. mongolica*, BEAST 2.2.1 was used to construct the Bayesian skyline plot (BSP) according to the

188   previous sequence mutation rate in Bayesian inference, and Tracer 1.5 was used to generate the BSP (Rambaut

189   & Drummond 2009).

190   Based on the assumption that sites with higher species distribution frequencies have lower migration

191   costs, the species distribution model was converted to a species habitat resistance model. We inverted the

192   species distribution model (1-SDM) for the last glacial maximum (LGM, based on MIROC-ESM Model) and

193   present (1970–2000) periods generated in MAXENT 3.4.1 to a "dispersal cost layer (resistance layer)";

194 Second, based on the least-cost path (LCP), the resistance layer of *A. mongolica* using SDMtoolbox 2.0

195 (Brown 2014) in ArcGIS 10.5 was used to create a cost distance raster for each sample locality, and the

196 corridor layers were established between every two localities based on the cost distance raster; Finally, all of

197 the pairwise corridor layers were summed as the eventual dispersal corridor for *A. mongolica* (Jiang et al.

198 2018). The data required for species distribution model of LGM and present periods were obtained from our

199 previous study (Ma et al. 2019).

Identification of potential protection areas

201 Comprehensive consideration of population genetic diversity, the suitable distribution habitats and the

202 Land use/cover were necessary for the scientific and effective conservation of species (Wei et al. 2014). The

203 dataset of Genetic diversity landscapes, Land Use/Cover Change (LUCC) and the species distributions

204 suitability of *A. mongolica* were employed to identify the potential conservation areas for *A. mongolica* (Wei

205 2020). Genetic diversity was mapped by genetic landscape GIS toolbox in ArcGIS 10.5 according to genetic

206 diversity values (H_d and R_d) of each studied population and interpolated by using an established inverse-

207 distance-weighted interpolation algorithm (Vandergast et al. 2011; Yu et al. 2014). LUCC (2018; 1 km×1 km)

208 in study area was obtained from the Data Center of Resources and Environment Sciences, Chinese Academy of

209 Sciences (http://www.resdc.cn/). According to the characteristics of suitable habitats of *A. mongolica* (Ma et al.

210 2015), a correlation value was assigned to each different land use type of LUCC, and the suitable habitats

211 (grassland, sand, bare rock, and others, etc.) were assigned values ranged from 0 to 1 (Table S3; Reng 2019).

212 Present potential suitable distribution data was obtained by species distribution models (MAXENT) from the

213 previous study (Ma et al. 2019). The dataset of Genetic diversity landscapes, LUCC and the potential

214 distribution were overlaid by using weight of 0.7, 0.2 and 0.1 in ArcGIS 10.5, and classified into three classes

215 using the natural discontinuity grading method, which were defined as the low potential protection areas,

216 medium potential protection areas and high potential protection areas (Wei 2020).

Environmental heterogeneity analysis of Geographically Weighted Regression

218 Geographically weighted regression model (GWR) was implemented to test the relationship between

219 genetic diversity values among *A. mongolica* populations and the environmental variables.

220 GWR was a spatial regression model that was an extension of traditional regression to predict, detect and

221   estimate spatial non-stationary coefficient for model variables (Daniel 2004), it can reflect the spatial

222   heterogeneity of variable distribution (Abebaw et al. 2021).

223   We used the genetic diversity values (Hd and Rd) of each population from both cpDNA and nDNA as the

224   response variable. As predictor variables, we choosed 8 environmental factors from species distribution model

225   and altitude data (Worldclim, http://www.worldclim.org), then we removed multicollinearity (variance

226   inflation factor>7.5) between variables independent variables, and finally we kept five factors which were

227   altitude, annual mean temperature (Bio1), isothermality (Bio3), temperature annual range (Bio7), and mean

228   temperature of driest quarter (Bio9). In order to obtain dependent variable, we used the fishing net tool

229   generate points with resolution of 2000 meters, and then extracted the dependent variable and independent

230   variable information in ArcGIS10.5. GWR4.08 (https://gwrtools.github.io/) software was used to finish

231   geographically weighted regression analysis, and ARCGIS10.5 was used to analyze the results visually.

232   **Results**

233   **The characteristics of cpDNA and nDNA sequences**

234   The cpDNA aligned sequences of *trn*L-*trn*F, *psb*K-*psb*I and *trn*V were 402 bp, 230 bp and 572 bp in

235   length respectively, and 1,204 bp in total. In the combined data (Table S1), we identified 17 different

236   haplotypes (H1-H17) and 24 polymorphic sites (18 substitutions and 6 indels). The aligned fragment ITS1-

237   ITS4 was 625 bp in length, with 9 nucleotide substitutions and 10 ribotypes (R1-R10) were defined (Table S2,

238   Fig. 1).

239   **Haplotype/ribotypes geographical distributions**

240   According to the cpDNA and nDNA dataset, the network analysis revealed two clades, the 9 haplotypes

241   (H1-H9) and 5 ribotypes (R1-R5) in the Clade 1 are from the populations of the Western group, and the

242   remaining 8 haplotypes (H10-H17) and 5 ribotypes (R6-R10) in the Clade 2 belong to the Eastern group, and

243   no genotype was shared between the two groups (Fig. 1b). The results of PCoA show that the haplotypes and

244   ribotypes mentioned above were independently clustered together, and the first three axes have explained a

245   cumulative percent variation of 81.19% and 80%, respectively (Fig. 2).

246   The Western group mainly distribute in Alxa Left Banner and Hexi Corridor. The most widespread

247   haplotype was H5, which shared by 4 populations and mainly distributed in Alxa Left Banner (ZQL, ZQK,

248 WSH and WSM), and there are 7 private haplotypes (1, 2, 4, and 6-9) in the Western group. The most common

249 ribotype was R3, which distributed among 5 populations, mainly from Hexi Corridor (YCQ) and Alxa Left

250 Banner (ZTM, ZQL, ZTL and ZQK; Table 1 and Fig. 1).

251 In the Eastern group, the most widespread haplotype was H12, which were detected among each of the 4

252 populations and distributed in Urad Zhongqi (WLJ), Yinshan Mountains (YMM and YMF), and Urad Houqi

253 (WLB); and 5 private haplotypes (10,13,14,16 and 17) were revealed. The most common ribotype was R8,

254 which distributed in Urad Zhongqi (WLJ), the Yinshan Mountains (YMF and YMM) and Urad Houqi (WLB);

255 R7, R9 and R10 were detected as 3 private ribotypes (Table 1 and Fig. 1).

256 **Population genetic diversity and genetic differentiation**

257 The cpDNA haplotype / nDNA ribotype diversity (Hd, Rd) among the 15 populations ranged from 0 to 0.

258 567 (cpDNA) and 0 to 0.676 (nDNA); Nucleotide diversity(π) ranged from 0 to 0.00962 (cpDNA) and 0 to 0.0

259 0922(nDNA), respectively (Table 1). According to both the cpDNA and nDNA datasets, the higher levels of

260 genetic variation were identified in the populations MZS (Mazong Mountains), ZQK (Southwestern Lang

261 Mountains) and WSH (Northern Helan Mountains) from the Western group, and the WLB (Urad Zhongqi) and

262 BTQ (Daqing Mountains) from the Eastern group (Fig. S2).

263 The H_T (cpDNA: 0.807 ± 0.050; nDNA: 0.815 ± 0.077) was much higher than H_S (cpDNA: 0.225 ±

264 0.053; nDNA: 0.212 ± 0.014) with both cpDNA and nDNA data, indicating considerable population

265 differentiation. A higher level of N_{ST} (0.858 / 0.829; $P < 0.05$) than G_{ST} (0.712 / 0.690; $P < 0.05$), indicated a

266 significant phylogeographical structure across the species range (Table 2). AMOVAs based on cpDNA and

267 nDNA sequences showed that 83.87% and 79.33% of the total variation primarily occurred among

268 populations, and 66.25% and 73.14% occurred among groups (Table 3). Results of PCoA showed that the

269 populations from the Western and Eastern groups based on cpDNA and nDNA were independently clustered

270 together, which revealed that the Western group was clearly distinct from the Eastern group (Fig. 2).

271 The genetic landscape shape analysis of *A. mongolica* showed significantly divergence in study

272 area, and signifcant genetic differentiation occurred between the Western and Eastern regions (Fig. 3).

273 Meanwhile, two strong genetic barriers between the Western and Eastern groups, with over 81% bootstrap

274 support, were detected in the Kubuqi and Ulanbuhe deserts based on Monmonier's maximum difference

275   algorithm (Fig. 1 c, d).

276   **Phylogenetic analysis and divergence time dating**

277   Similar gene genealogies of genotypes in the BI trees and ML trees based on cpDNA and nDNA were

278   obtained (Figs. 4 and S1). The nine haplotypes (1-9) and five ribotypes (1-5) corresponded to the Western

279   group, while the remaining eight haplotypes (10-17) and five ribotypes (6-10) corresponded to the Eastern

280   group (Figs. 4and S1).

281   BEAST analysis showed that the cpDNA and nDNA divergence between the two groups of *A. mongolica*

282   were 0.28 million years ago (Ma) and 0.31 Ma (Fig. 4). The timing of divergence for two groups are consistent

283   with the middle Pleistocene during a period of strong expansion of aridification in northwestern China (Comes

284   & Kadereit 1998).

285   **Demographic history analysis**

286   The parameters of Tajima's *D* and Fu's *Fs* for the overall population and the two groups were not

287   significant and were positive, indicating that *A. mongolica* has not experienced a recent expansion (Table 4).

288   Despite this, post-glacial range expansion indicated that the overall population as well as the two groups had

289   evidence of recent expansion through the unimodal mismatch distribution curves of nDNA (Fig. 5). This

290   conclusion was also supported by the *SSD* and H_{rag} which were not significant ($p > 0.05$) in cpDNA and

291   nDNA for the overall populations or the two groups in cpDNA and nDNA (Table 4). However, the Bayesian

292   skyline plot summarize instantaneous estimates of effective population size based on cpDNA for the total

293   populations, the Eastern group, and the Western group, and showed recent population decline for *A.*

294   *mongolica*. A slight population decline was found for the Eastern group and Western group, and a contraction

295   followed by a small amount of population growth was found for the total population according to the nDNA

296   (Fig. 6).

297   In the LGM based on MIROC-ESM climate model, the Lang Mountains, the edge of Tennger deserts, and

298   the Hexi Corridor are the important corridors for the dispersal of *A. mongolica* (Fig. 7a). In the present period,

299   the dispersal corridors of *A. mongolica* were similar to the model, but the edges of the Ulanbuhe deserts was

300   identified as the most important migration corridors for the species (Fig. 7b).

301   **Identification of potential protection areas**

302   Considering the overall genetic diversity, distribution suitability of *A. mongolica* and Land use/Cover in

303   northwestern China, high potential protection areas were identified in the south slopes of the Lang Mountains

304   and the northern Helan Mountains (Fig. 8). These two areas had high genetic diversity based on cpDNA and

305   nDNA (H_d and R_d >0.70), high distribution suitability as determined by MAXENT (>0.80), and more natural

306   habitats with little human activity (including sand, bare rock and grasslands).

307   **Environmental heterogeneity analysis**

308   The spatial distribution of regression coefficients of different environmental factors was significantly

309   different across in the study area, indicating spatial heterogeneity between different environmental factors and

310   genetic diversity of *A. mongolica* (Figs. S4, S5). In the Eastern group region, genetic diversity was positively

311   correlated with altitude, Bio1, Bio3, and Bio7, but was negatively correlated with Bio9. In the Western group

312   region, genetic diversity was positively correlated with Bio1, Bio3, Bio7, and Bio9, but was negatively

313   correlated with altitude (Figs. S4, S5).

314   **Discussion**

315   **Allopatric divergence between the Western and Eastern regions**

316   The strong phylogeographic structure of *A. mongolica* indicated that separate and isolated lineages

317   occupy the different geographic regions (Fig. 1). Both the chloroplast and nuclear phylogenetic analyses

318   showed two distinct lineages distributed in the Western and Eastern regions (Fig. 1). The network and PCoA

319   analysis also indicated similar results: with 9 haplotypes and 5 ribotypes from the Western region clustered

320   together and apparently separated from the other 7 haplotypes and 5 ribotypes distributed in the Eastern region

321   (Figs. 1, 2).

322   No incomplete genealogical sorting was detected within the distribution range of *A. mongolica*, indicating

323   that the polymorphism of molecular markers used in this study were relatively appropriate. By contrast, we

324   previously identified a large number of shared haplotypes (H2, 4, 6) between the Western and Eastern groups

325   based on two cpDNA sequences, and phylogenetic analysis also revealed incomplete genealogical sorting (Ma

326   et al. 2019). More than 75% of the observed differences were revealed between the two regions. The

327   significant divergence among the Western and Eastern population groups was estimated to occur during mid-

328   Pleistocene (Figs. 3, 4). We therefore hypothesize that there is a close link between the increased aridification

329   and desertification during the Quaternary glacial periods and intraspecific divergence of *A. mongolica*.

330   The mountains and deserts in the distribution areas of *A. mongolica* have leaded to distinct patterns of

331   genetic isolation between 15 populations (Fig. 1), and these arid mountains slope and sand likely acted as

332   geographical barriers and significantly affected interspecific differentiation of the desert land vegetation in

333   northwestern China, such as *Euphrates poplar*, *Gymnocarpos przewalskii*, *Reaumuria soongarica* and *Lycium*

334   *ruthenicum* (Jiang et al. 2018; Shi et al. 2020; Wang et al. 2021). For *A. mongolica*, the pre-existence and rapid

335   expansions of the Ulanbuhe and Kubuqi deserts during Pleistocene (Guan et al. 2011), increased aridification

336   and likely create potential barriers to gene flow between populations in the eastern and western regions. The

337   two deserts were inferred as the genetic barriers for the two groups of *A. mongolica* (Fig. 1c, d).The sampled

338   populations located in the different secondary valleys of the Yinshan Mountains, and the uplifts and eastward

339   extensions of these valleys since the Quaternary period (Lee & Zhang 2010), which likely further isolated the

340   Eastern populations from the populations in the western regions. Moreover, in recent decades, with the

341   intensification of drought and desertification in the arid northwestern China added the change of Land

342   use/Cover, *A. mongolica* and its natural habitats have been greatly disturbed and destroyed (Duan et al. 2020;

343   Ma et al. 2015), the arid landscape fragmentation have caused the fragmented distributions and populations

344   isolation, which hence restricted movement of pollen and seed of *A. mongolica*. Thus, the fragmentation of

345   arid landscape was the potential barriers limiting gene flow, and it probably increased genetic differentiation

346   among different populations. Furthermore, long-term isolation among plant populations in heterogeneous

347   habitats can gradually lead to local adaptive differentiation and eventually create genetic heterogeneity across

348   different landscapes (Chen et al. 2020; Jiang et al. 2018). For *A. mongolica*, the typical mountain basin–deserts

349   isolation patterns between the eastern and western sampled areas, the arid landscape fragmentation and the

350   corresponding environmental heterogeneity (Figs S4, S5) have combined to promote differentiation between

351   the Western-Eastern regions through local environmental adaptation.

352   **Pleistocene population history of *A. mongolica***

353   Climatic oscillations in the Quaternary period have caused range shifts for many xerophytic plants in the

354   arid regions of northwestern China (Zhang et al. 2016). During glacial periods, most desert plants reduced their

355   distributions due to the unusually cold-dry climates, and had recolonized areas that are more suitable at the end

356   of glacial periods (Sun et al. 2020; Zhao et al. 2019a). The unimodal mismatch distribution curves based on

357   ITS sequences along with significant SSD and H_{rag} values all demonstrate that *A. mongolica* and the two

358   population groups underwent post-glacial range expansions (Table 4 and Fig. 5). Considering the phylogenetic

359   network according to cpDNA and nDNA datasets, genotypes H5/R3 and H12/R8 had high frequencies in the

360   Western and Eastern groups, and "star-like'' clusters originating from these genotypes mentioned were found

361   (Fig. 1), also reflecting the historical signature of post-glacial expansions in the two regions. Plants in rapidly

362   colonized regions generally possess low levels of genetic variation (Hewitt 2000). For the Western group, the

363   widespread distribution of H5 and R3 in the Alxa Left Banner (ZQK, WSH, WSM, ZQL, ZTM and ZTL) and

364   the eastern Hexi Corridor (YCQ) supports the hypothesis that *A. mongolica* expanded westwards along the

365   margins of the Ulanbuhe and Tengger deserts to the Hexi Corridor. Similarly, the demographic pattern of the

366   widespread genotypes H12 and R8 revealed an expansionary trajectory of the Eastern group along the eastern

367   edge of the Lang Mountains (Fig. 1). Moreover, the Bayesian skyline plot of species as determined by nDNA

368   also identified a recent population expansion. Compared with cpDNA, the genetic information of nDNA can

369   reflect the recent gene status. The low replacement rate of nDNA reduces the anti-mutation and parallel

370   mutation, including the fixed differentiation information of a single locus and low convergence data to lead to

371   better statistical results (Harris & Hey 1999). The recent population expansion revealed by nDNA was likely

372   associated with the warm and humid climate of the post-glacial period (Su & Zhang 2013).

373   The Western and Eastern groups of *A. mongolica* experienced rapid range expansions during the Late

374   Pleistocene-Early Holocene (Fig. 5). During the Late Pleistocene, the continued progress of aridification in

375   northwestern China contributed to desert expansions, which in some cases may have provided a broader

376   appropriate habitats (such as at the edges of deserts and arid piedmont grassland) for expansions of arid land

377   plants (Ma et al. 2012). The inferred post-glacial expansion tracks along desert margins have been revealed in

378   some desert shrubs in northwestern China, including *Reaumuria soongarica*, *Nitraria tangutorum* and

379   *Gymnocarpos przewalskii* (Jiang et al. 2018; Shi et al. 2020; Sun et al. 2020). For *A. mongolica*, the Western

380   and Eastern groups were also found to have experienced significant outward expansions since the LGM, likely

381   along the edges of the Ulanbuhe and Tengger deserts and Lang Mountains, respectively (Fig. 7). The warmer

382   climates during the early Holocene facilitated snow melting on many of the peaks of the middle and south of

383   the Qilian, Helan and Lang Mountains, a greater amount of runoff from melting snow and glacial ice resulted,

384   making the local habitats of *A. mongolica*, such as the edges of the Ulanbuhe and Tengger deserts and the Lang

385   Mountains wetter (Shi et al. 2003; Yang et al. 2011). The wet conditions favored the eastward migration of the

386   eastern populations along the Lang Mountains and the westward migration of the western populations along

387   the edges of the Ulanbuhe and Tengger deserts towards the Hexi Corridor.

388   **Conservation implications**

389   Habitat fragmentation and reduction, such as increasing arable land area and urbanization, accompanied

390   by low genetic variation, could ultimately increase the chance of genetic drift and inbreeding within

391   populations. In this study, the arid landscape fragmentation led to the low levels of genetic variation within

392   populations in the YMM (mid-levels of the Yinshan Mountains), YMF (foot of the Yinshan Mountains) and

393   WLJ (Urad Zhongqi) in the Eastern group, and in the WSH (northern Helan Mountains), WSM (Wusutu town

394   of Alxa Left Banner), and ZQL (Southern Helan Mountains) in the Western group (Fig. 1 and Table 1). An

395   integrated assessment and identification of conservation centers is a more effective way to optimize

396   conservation strategies for *A. mongolica* (Zhang et al. 2018). Overall, conservation populations with high

397   levels of genetic diversity facilitate the maintenance of evolutionary potential and the potential for adaptation

398   to future environment change (Zhang et al. 2020). However, it is important to consider habitat suitability and

399   the current and future Land use/Cover during establishment of potential protected areas. Because of this, the

400   south slopes of the Lang Mountains and the northern Helan Mountains were identified as potential important

401   protected areas, as they would be beneficial for accumulation of genetic variation, and generation of genetic

402   mutations, ultimately increasing the evolutionary potential of *A. mongolica*. For these two protected areas, it is

403   recommended that suitable strategies be adopted for in situ conservation, considering the habitats in both areas

404   are relatively intact.

405   **Conclusions**

406   This study suggests a combination of climatic fluctuations and aridification during the Quaternary period

407   combined with significant environmental heterogeneity in the Western and Eastern groups to play an important

408   role in the phylogeographic structure and genetic diversity of *A. mongolica*. For this species, the enhanced

409   aridity along with expansions of the Ulanbuhe and Tengger deserts, and the eastward extension of the Yinshan

410 Mountains since the Quaternary period likely interrupted gene flow, triggering the current divergence in the

411 two allopatric regions. The south slopes of the Lang Mountains and northern Helan Mountains were identified

412 as having the most potential to be important protected areas for *A. mongolica* which is meaningful for the

413 management of this endangered species. Future experiments will focus on how to better quantitatively describe

414 gene flow.

415 **Conflicts of Interest:** The author declares no conflict of interest.

416

417 **Data Archiving Statement**

418 All haplotype and ribotypes sequences gathered in this study have been deposited in GenBank (*trnL-trnF*,

419 MK248504-MK248515; *psb*K-*psb*I, MK248492-MK248503; *trn*V, MT387487-MT387503; ITS1- ITS4,

420 MT397024-MT397033). The accession numbers of the outgroup sequences for haplotype and ribotypes of

421 *Amygdalus davidiana* are MK798145.1 and MH711207.1. The accession numbers of the outgroup sequences

422 for haplotype and ribotypes of *Prunus triloba* are MK790138.1 and EU669088.1, respectively.

423

424 **References**

425 Abebaw TS, Tefera FD, Tigabu YS, Asmamaw AA, and Alemu GK. 2021. Spatial distribution and geographical
426   heterogeneity factors associated with poor consumption of foods rich in vitamin A among children age 6-
427   23 months in Ethiopia: Geographical weighted regression analysis. *Plos One* 16:e0252639.
428   DOI:10.1371/journal.pone.0252639

429 Arbogast BS. 2001. Phylogeography: The History and Formation of Species. *Integrative and Comparative Biology*
430   41:134-135. DOI:10.1093/icb/41.1.134

431 Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, and V.Alekseyenko A. 2012. Improving the Accuracy of
432   Demographic and Molecular Clock Model Comparison While Accommodating Phylogenetic Uncertainty.
433   *Molecular Biology & Evolution* 12:2157-2167. DOI:10.1093/molbev/mss084

434 Bandelt HJ, Forster P, and Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular*
435   *Biology and Evolution* 16:37-48. DOI:10.1093/oxfordjournals.molbev.a026036

436 Brown JL. 2014. SDM toolbox: a python‑based GIS toolkit for landscape genetic, biogeographic and species
437   distribution model analyses. *Methods in Ecology and Evolution* 5:694-700. DOI:10.1111/2041-210X.12200

438 Chen Q, Yin Y, Zhao R, Yang Y, Teixeira DSJA, and Yu X. 2020. Incorporating Local Adaptation Into Species Distribution
439   Modeling of Paeonia mairei, an Endemic Plant to China. *Frontiers in Plant Science* 10:1717.
440   DOI:10.3389/fpls.2019.01717

441 Comes HP, and Kadereit JW. 1998. The effect of quaternary climatic changes on plant distribution and evolution.
442   *Plant Science* 3:432-438. DOI:10.1016/S1360-1385(98)01327-2

443 Daniel PM. 2004. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. *American*

444   *Journal of Agricultural Economics* 86:554-556. DOI:10.1111/j.0002-9092.2004.600_2.x

445   Delplancke M, Yazbek M, Arrigo N, Espíndola A, Joly H, and Alvarez N. 2016. Combining conservative and variable
446   markers to infer the evolutionary history of Prunus subgen. Amygdalus s.l. under domestication. *Genetic*
447   *Resources&Crop Evolution* 63:221-234. DOI:10.1007/s10722-015-0242-6

448   Doyle JJ. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19:11-
449   15. DOI:

450   Drummond AJ, and Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *Bmc Evolutionary*
451   *Biology* 7:214. DOI:10.1186/1471-2148-7-214

452   Duan Y, Wang J, Wang C, Wang H, and Du Z. 2020. Analysis on the potential suitable areas of four species of the
453   *subgen. Amygdalus* in arid Northwest China under future climate change. *Chinese Journal of Ecology*
454   39:2193-2204. DOI:10.13292/j.1000-4890.202007.023

455   Excoffier L, Laval G, and Schneider S. 2005. Arlequin (version 3.0): An integrated software package for population
456   genetics data analysis. *Evolutionary Bioinformatics Online* 1:47-50. DOI:10.1143/JJAP.34.L418

457   Excoffier L, and Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics
458   analyses under Linux and Windows. *Molecular Ecology Resources* 10:564-567. DOI:10.1111/j.1755-
459   0998.2010.02847.x

460   Fu LG, and Chin CM. 1992. *China plant red data book : rare and endangered plants*: China plant red data book : rare
461   and endangered plants.

462   Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background
463   selection. *Genetics* 147:915-925. DOI:10.1017/S0016672397002966

464   Gao XY, Meng HH, and Zhang ML. 2014. Diversification and vicariance of desert plants: Evidence inferred from
465   chloroplast DNA sequence variation of Lagochilus ilicifolius (Lamiaceae). *Biochemical Systematics & Ecology*
466   55:93-100. DOI:10.1016/j.bse.2014.03.002

467   Guan Qy, Pan Bt, Li N, Zhang Jd, and Xue Lj. 2011. Timing and significance of the initiation of present day deserts in
468   the northeastern Hexi Corridor, China. *Palaeogeography Palaeoclimatology Palaeoecology* 306:70-74.
469   DOI:10.1016/j.palaeo.2011.03.029

470   Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows
471   95/98/NT. *Nucleic acids symposium series*:95-98. DOI:10.1021/bk-1999-0734.ch008

472   Harpending HC. 1994. Signature of Ancient Population Growth in a Low-Resolution Mitochondrial DNA Mismatch
473   Distribution. *Human Biology* 66:591-600. DOI:10.1038/hdy.1994.122

474   Harris EE, and Hey J. 1999. X chromosome evidence for ancient human histories. *Proceedings of the National*
475   *Academy of Sciencesof the United States of America* 96:3320-3324. DOI:10.1073/pnas.96.6.3320

476   Harrison SP, G. Yu HT, and Prentice IC. 2001. Palaeovegetation (communications arising): diversity of temperate
477   plants in East Asia. *Nature* 413:129-130. DOI:10.1038/35093166

478   Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405:907-913. DOI:10.1038/35016000

479   Hewitt GM. 1996. Some genetic consequences of ice ages, and their role, in divergence and speciation. *Biological*
480   *Journal of the Linnean Society* 58:247-276. DOI:10.1111/j.1095-8312.1996.tb01434.x

481   Jia SW, and Zhang ML. 2021. Introgression of phylogeography lineages of Convolvulus gortschakovii
482   (Convolvulaceae) in the northwest China. *Plant Systematics and Evolution* 307:0378-2697.
483   DOI:10.1007/s00606-020-01734-z

484   Jiang XL, An M, Zheng SS, Deng M, and Su ZH. 2018. Geographical isolation and environmental heterogeneity

485   contribute to the spatial genetic patterns of Quercus kerrii (Fagaceae). *Heredity* 120:219-233.
486    DOI:10.1038/s41437-017-0012-7

487 Lee H, and Zhang D. 2010. Natural disasters in northwestern China, AD 1270–1949. *Climate Research* 41:245-257.
488    DOI:10.3354/cr00858

489 Li Y, Song F, Zhang XN, and Lv GH. 2019. Phylogeography suggest the Yili Valley being the glacial refuge of the genus
490    Ixiolirion (Amaryllidaceae) in China. *Systematics & Biodiversity* 17:1-17.
491    DOI:10.1080/14772000.2019.1612966

492 Librado P, and Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.
493    *Bioinformatics* 25:1451-1452. DOI:10.1093/bioinformatics/btp187

494 Ma SM, Nie YB, Duan X, Yu CS, and Wang RX. 2015. The potential distribution and population protection priority of
495    Amygdalusmongolica. *ACTA ECOLOGICA SINICA* 35:2960-2966. DOI:10.5846/stxb201308292173

496 Ma SM, Nie YB, Jiang XL, Xu Z, and Ji WQ. 2019. Genetic structure of the endangered, relict shrub Amygdalus
497    mongolica (Rosaceae) in arid northwest China. *Australian Journal of Botany* 67:128-139.
498    DOI:10.1071/BT18188

499 Ma SM, Zhang ML, and Sanderson SC. 2012. Phylogeography of the rare *Gymnocarpos przewalskii* (Caryophyllaceae):
500    indications of multiple glacial refugia in north-western China. *Australian Journal of Botany* 60:20-31.
501    DOI:10.1071/BT11055

502 Manni F, Guérard E, and Heyer E. 2004. Geographic Patterns of (Genetic, Morphologic, Linguistic) Variation: How
503    Barriers Can Be Detected by Using Monmonier's Algorithm. *Human Biology* 76:173-190.
504    DOI:10.1353/hub.2004.0034

505 Meng HH, and Zhang ML. 2011. Phylogeography of Lagochilus ilicifolius (Lamiaceae) in relation to Quaternary
506    climatic oscillation and aridification in northern China. *Biochemical Systematics and Ecology* 39:787-796.
507    DOI:10.1016/j.bse.2011.07.015

508 Miller MP. 2005. ALLELES IN SPACE (AIS): computer software for the joint analysis of interindividual spatial and
509    genetic information. *Journal of Heredity* 96:722-724. DOI:10.1093/jhered/esi119

510 Ni J, Harrison SP, Prentice IC, Kutzbach JE, and Sitch S. 2006. Impact of climate variability on present and Holocene
511    vegetation: A model-based study. *Ecological Modelling* 191:469-486. DOI:10.1016/j.ecolmodel.2005.05.019

512 Oksanen J, Kindt R, Legendre P, O'Hara B, Gavin L. Simpson, Solymos P, Stevens MHH, and Wagner H. 2007. The
513    vegan package. *Community Ecology Package* 10:631–637. DOI:

514 Pons O, and Petit R. 1996. Measwring and testing genetic differentiation with ordered versus unordered alleles.
515    *Genetics* 144:1237-1245. DOI:10.1016/S1050-3862(96)00162-3

516 Posada D, and Crandall KA. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14:817–818.
517    DOI:10.1093/bioinformatics/14.9.817

518 Qiu Y, Lu Q, Zhang Y, and Cao Y. 2017. Phylogeography of East Asia's Tertiary relict plants: current progress and
519    future prospects. *Biodiversity Science* 25:136-146. DOI:10.17520/biods.2016292

520 Rambaut A, and Drummond AJ. 2009. Tracer Version 1.5.0. *Available at http://beast.bio.ed.ac.uk/software/tracer/*.

521 Ren J. 2019, Integrating phylogeography and landscape genetics reveals formation mechanism of genetic
522    variation and its spatial distribution pattern for stipa breviflora. D. –Neimenggu Univ.

523 Riccioni C, Rubini A, Türkoğlu A, Belfiori B, and Paolocci F. 2019. Ribosomal DNA polymorphisms reveal genetic
524    structure and a phylogeographic pattern in the Burgundy truffle Tuber aestivum Vittad. *Mycologia* 111:1-
525    14. DOI:10.1080/00275514.2018.1543508

526   Richardson JE, Pennington RT, Pennington TD, and Hollingsworth PM. 2001. Richardson JE, Pennington RT,
527   Pennington TD, Hollingsworth PM. Rapid diversification of a species-rich genus of neotropical rain forest
528   trees. *Science* 293:2242-2245. DOI:10.1126/science.1061421
529   Rogers, and Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences.
530   *Molecular Biology & Evolution* 9:552-569. DOI:10.1093/oxfordjournals.molbev.a040727
531   Shaw J, Lickey EB, Schilling E, and Small RL. 2007. Comparison of whole chloroplast genome sequences to choose
532   noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal*
533   *of Botany* 94:257-288. DOI:10.3732/ajb.94.3.275
534   Shi XJ, and Zhang ML. 2015. Phylogeographical structure inferred from cpDNA sequence variation of Zygophyllum
535   xanthoxylon across north-west China. *Journal of Plant Research* 128:269-282. DOI:10.1007/s10265-014-
536   0699-y
537   Shi Y, Chen Y, Li D, Zhang G, Ding Y, Ruji H, and Kang E. 2003. Disscussion on the present climate changefrom warm-
538   dry to warm-wet in northwest China. *Quaternary Sciences* 23:152-164. DOI:
539   Shi Y, Yan X, Yin H, Qian C, Fan X, Yin X, Chang Y, Zhang C, and Ma X. 2020. Divergence and hybridization in the desert
540   plant Reaumuria soongarica. *Journal of Systematics and Evolution* 58:159-173. DOI:10.1111/jse.12490
541   Shi YF. 2006. *The Quaternary Glaciations and Environmental Variations in China*.
542   Simmons M, and Ochoterena H. 2000. Gaps as Characters in Sequence-Based Phylogenetic Analyses. *Systematic*
543   *Biology* 49:369-381. DOI:10.1080/10635159950173889
544   Stephen M, and Donnelly P. 2003. A Comparison of Bayesian Methods for Haplotype Reconstruction from Population
545   Genotype Data. *American Journal of Human Genetics* 73:1162-1169. DOI:10.1086/379378
546   Su Z, and Zhang M. 2013. Evolutionary response to Quaternary climate aridification and oscillations in north-western
547   China revealed by chloroplast phylogeography of the desert shrub *Nitraria sphaerocarpa* (Nitrariaceae).
548   *Biological Journal of the Linnean Society* 109:757-770. DOI:10.1111/bij.12088
549   Sudhir K, Glen S, Michael L, Christina K, and Koichiro T. 2018. MEGA X: Molecular Evolutionary Genetics Analysis
550   across Computing Platforms. *Molecular Biology and Evolution* 35:1547–1549. DOI:10.1093/molbev/msy096
551   Sun L, Liu G, Lu Y, Zhang B, and Zhang G. 2020. Molecular data and ecological niche modelling reveal the
552   phylogeographic pattern of the widespread shrub Tamarix chinensis Lour. (Tamaricaceae) in China. *Kew*
553   *Bulletin* 75:41. DOI:10.1007/s12225-020-09899-z
554   Tajima F. 1989. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics*
555   123:585-595. DOI:10.1093/genetics/123.3.585
556   Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, and Higgins DG. 1997. The CLUSTAL_X windows interface:
557   flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*
558   25:4876-4882. DOI:10.1093/nar/25.24.4876
559   Vandergast AG, Perry WM, Lugo RV, and Hathaway SA. 2011. Genetic landscapes GIS Toolbox: tools to map patterns
560   of genetic divergence and diversity. *Molecular Ecology Resources* 11:158-161. DOI:10.1111/j.1755-
561   0998.2010.02904.x
562   Wang CC, Ma SM, Sun FF, Wei B, and Nie YB. 2021. Spatial genetic patterns of the medicinal and edible shrub Lycium
563   ruthenicum (Solanaceae) in arid Xinjiang, China. *Tree Genetics & Genomes* 17:1-13. DOI:10.1007/s11295-
564   020-01488-2
565   Wei B. 2020. Genetic characteristics and conservation strategies of Haloxylon ammodendron populations in
566   desert. D. –Shihezi Univ.

567   Wei F, Nie Y, Miao H, Lu H, and Hu Y. 2014. Advancements of the researches on biodiversity loss mechanisms. *Chinese*
568   *Science Bulletin* 59:430-437. DOI:10.1360/972013-557

569   White T, Bruns T, Lee S, Taylor F, White T, Lee SH, Taylor L, and Shawetaylor J. 1990. Amplification and direct
570   sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols*:315-322. DOI:10.1016/B978-0-
571   12-372180-8.50042-1

572   Wickham H. 2016. About the ggplot2 Package. *Journal of Applied & Computational Mathematics* 5:1-12.
573   DOI:10.4172/2168-9679.1000321

574   Wolfe KH, Li W-H, and Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial,
575   chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of*
576   *America* 84:9054-9058. DOI:10.1073/pnas.84.24.9054

577   Xu Z, and Zhang ML. 2015. Phylogeography of the Arid Shrub Atraphaxis frutescens (Polygonaceae) in Northwestern
578   China: Evidence From cpDNA Sequences. *Journal of Heredity* 106:184-195. DOI:10.1093/jhered/esu078

579   Yang XL, Ding WK, Yang JH, and Zhou WX. 2011. Change characteristics and 5 stations contrastive analysis of climate
580   in eastern Hexi corridor in recent 50 years. *Agricultural Research in the Arid Areas* 29:259-268. DOI:1000-
581   7601(2011)05-0259-10

582   Yu G, Chen X, Ni J, Cheddadi R, Guiot J, Han H, Harrison S, Huang C, Ke M, and Kong Z. 2000. Palaeovegetation of
583   China: a pollen data-based synthesis for the mid-Holocene and last glacial maximum. *Journal of*
584   *Biogeography* 27:635-664. DOI:10.1046/j.1365-2699.2000.00431.x

585   Yu H, Zhang Y, Gao J, and Qi W. 2014. Visualizing Patterns of Genetic Landscapes and Species Distribution of Taxus
586   wallichiana (Taxaceae), Based on GIS and Ecological Niche Models. *Journal of Resources and Ecology* 5:193-
587   202. DOI:10.5814/j.issn.1674-764x.2014.03.001

588   Zeng YF, Zhang JG, Bawerjan A, Wang WT, and Jia ZQ. 2018. Phylogeographic patterns of the desert poplar in
589   Northwest China shaped by both geology and climatic oscillations. *Bmc Evolutionary Biology* 18:1-14.
590   DOI:10.1186/s12862-018-1194-1

591   Zhang HX, Li HY, and Li YX. 2018. Identifying evolutionarily significant units for conservation of the endangered Malus
592   sieversii using genome‐wide RADseq data. *Nordic Journal of Botany* 36. DOI:10.1111/njb.01733

593   Zhang HX, Wang Q, and Jia SW. 2020. Genomic phylogeography of Gymnocarpos przewalskii (Caryophyllaceae):
594   insights into habitat fragmentation in arid Northwestern China. *Diversity* 12:335. DOI:10.3390/d12090335

595   Zhang HX, and Zhang ML. 2012. Genetic structure of the Delphinium naviculare species group tracks Pleistocene
596   climatic oscillations in the Tianshan Mountains, arid Central Asia. *Palaeogeography, Palaeoclimatology,*
597   *Palaeoecology* 353-355:93-103. DOI:10.1016/j.palaeo.2012.07.013

598   Zhang HX, Zhang ML, and Sanderson SC. 2016. Spatial genetic structure of forest and xerophytic plant species in arid
599   Eastern Central Asia: insights from comparative phylogeography and ecological niche modelling. *Biological*
600   *Journal of the Linnean Society* 120:612-625. DOI:10.1111/bij.12903

601   Zhang HX, Zhang ML, Sanderson SC, and David C. 2013. Retreating or Standing: Responses of Forest Species and
602   Steppe Species to Climate Change in Arid Eastern Central Asia. *Plos One* 8:e61954.
603   DOI:10.1371/journal.pone.0061954

604   Zhang Ly, and Jiang Zl. 1992. The genesis of the arid climate of Northwest China. *Arid Land Geography* 1992:1-12.
605   DOI:10.13826/j.cnki.cn65-1103/x/1992/02/001

606   Zhao Y, Pan B, Zhang M, and Chiang TY. 2019a. Phylogeography and conservation genetics of the endangered
607   Tugarinovia mongolica (Asteraceae) from Inner Mongolia, Northwest China. *Plos One* 14:e0211696.

608   DOI:10.1371/journal.pone.0211696

609   Zhao Y, Zhang H, Pan B, and Zhang M. 2019b. Intraspecific divergences and phylogeography of *Panzerina lanata*

610   (Lamiaceae) in northwest China. *PeerJ* 7:e6264. DOI:10.7717/peerj.6264

611

PeerJ

Figure 1

Map of the sampling sites, geographical distribution of the chlorotypes and ribotypes, the phylogenetic network and genetic barriers among populations of *Amygdalus mongolica*.

a Sampling localities and geographic distribution of 17 cpDNA haplotypes (labelled as H1–H17) and 10 nDNA ribotypes (labelled as R1–R10), identified from 15 populations of *A. mongolica* in Northwest China. Pie graphs reflect the frequency of each haplotype (the black pie) and each ribotype (the red pie), the population codes listed in Table 1. **b** The median-joining network for the 17 haplotypes and 10 ribotypes. The sizes of the circles in the network are proportional to haplotype / ribotype frequencies. Branch lengths are roughly proportional to the number of mutation steps between haplotypes / ribotypes and nodes. **c, d** Genetic barriers to chlorotypes and ribotypes between different sampling areas, respectively.

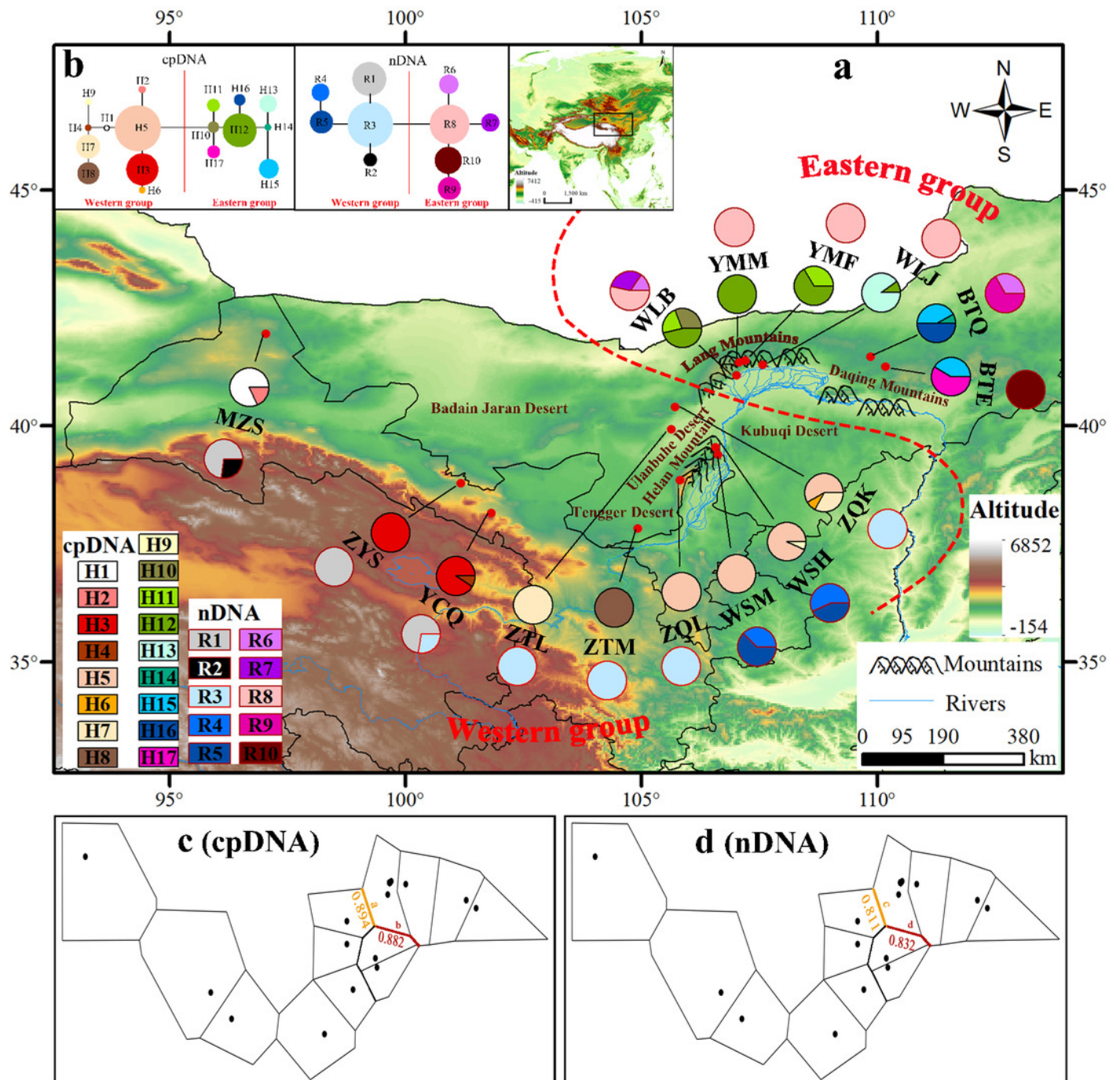



Figure 2

Plots of the first three coordinates of the principal coordinates analysis (PCoA) at a population level based on the cpDNA and nDNA pairwise differentiation matrix for *Amygdalus mongolica.*

Colors of dots represent the individuals in the Western and Eastern groups (as in Figure 1).

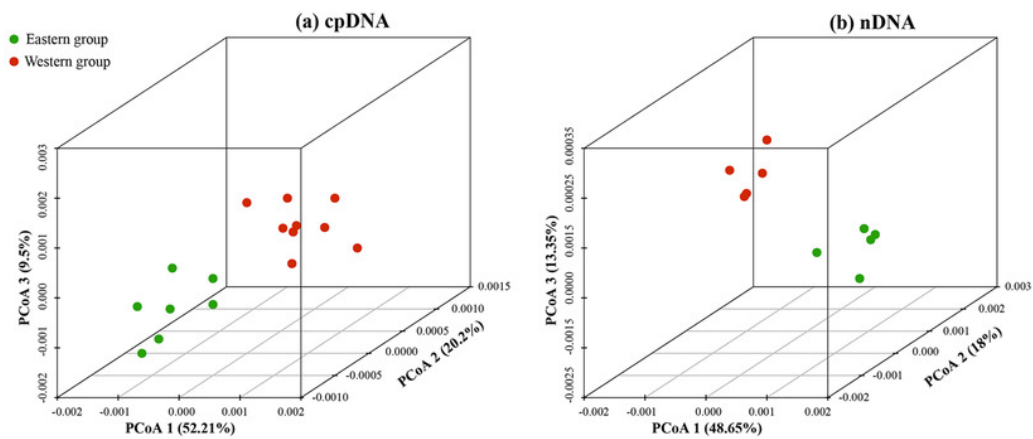



Figure 3

Spatial genetic landscapes constructed from the cpDNA sequences and nDNA sequences across the total distributions of *Amygdalus mongolica*.

The abscissae and ordinates correspond to geographical coordinates covering the entire distributional populations, and the vertical axes represent genetic distances.

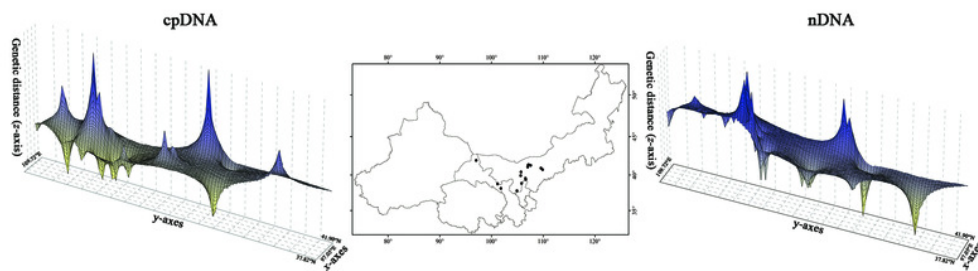



Figure 4

Bayesian divergence time estimates of *Amygdalus mongolica* based on 17 cpDNA haplotypes (left) and 10 nDNA ribotypes (right).

The values above branches represent the posterior probabilities above 0.65 and the divergence time are show below branches.

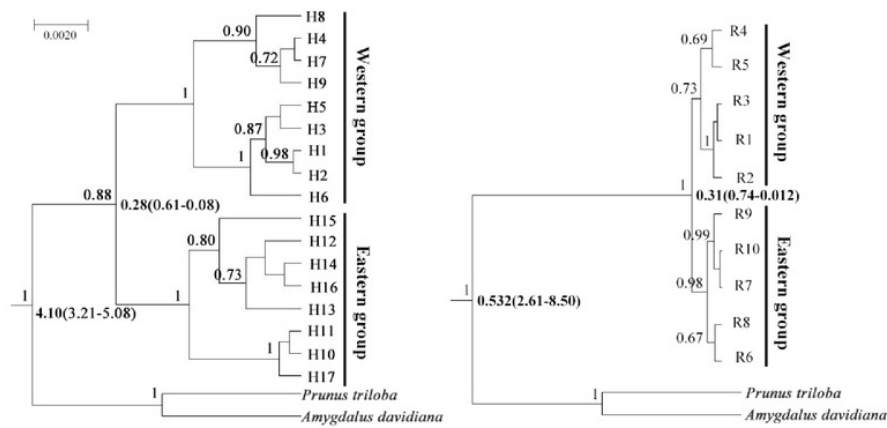



Figure 5

Mismatch distribution analysis based on the cpDNA and nDNA sequences for the total populations, as well as populations in the Eastern and Western groups (i.e., the grouping of populations consistent with Figure 1).

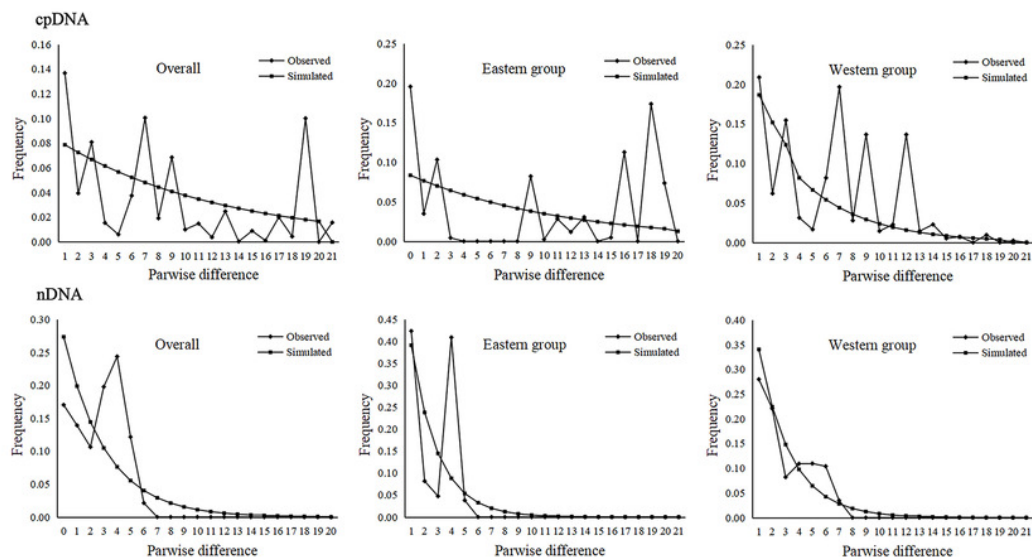



Figure 6

Bayesian skyline plot (BSP) estimating the time of *Amygdalus mongolica* population expansion based on the cpDNA and nDNA.

The solid line indicates the median values, and the area between both purple lines represents the boundary of the 95% central posterior density interval. X-axis: time, years before present (Ma); Y-axis, effective population size (*N*e, the product of effective population size and generation length).

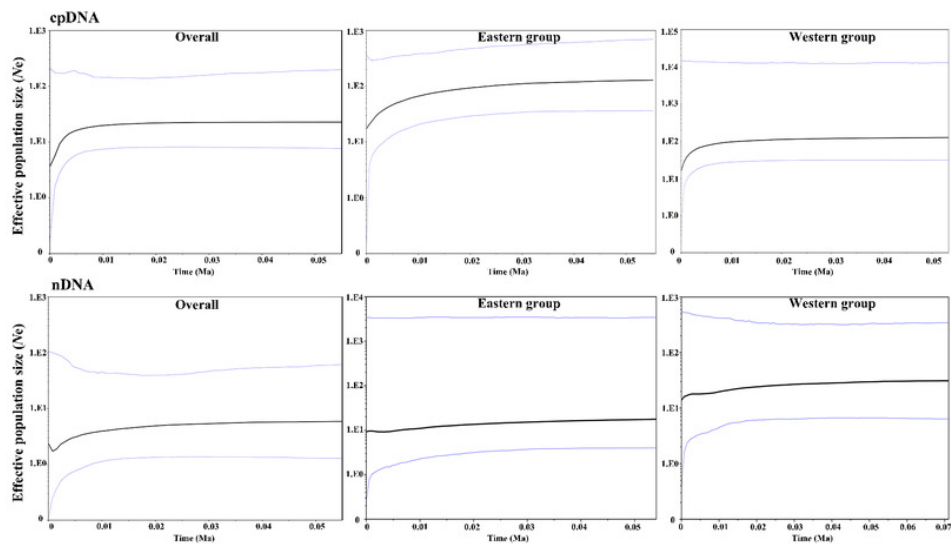



Figure 7

Dispersal corridors of *Amygdalus mongolica* during (a) the last glacial maximum (LGM) period and (b) the present.

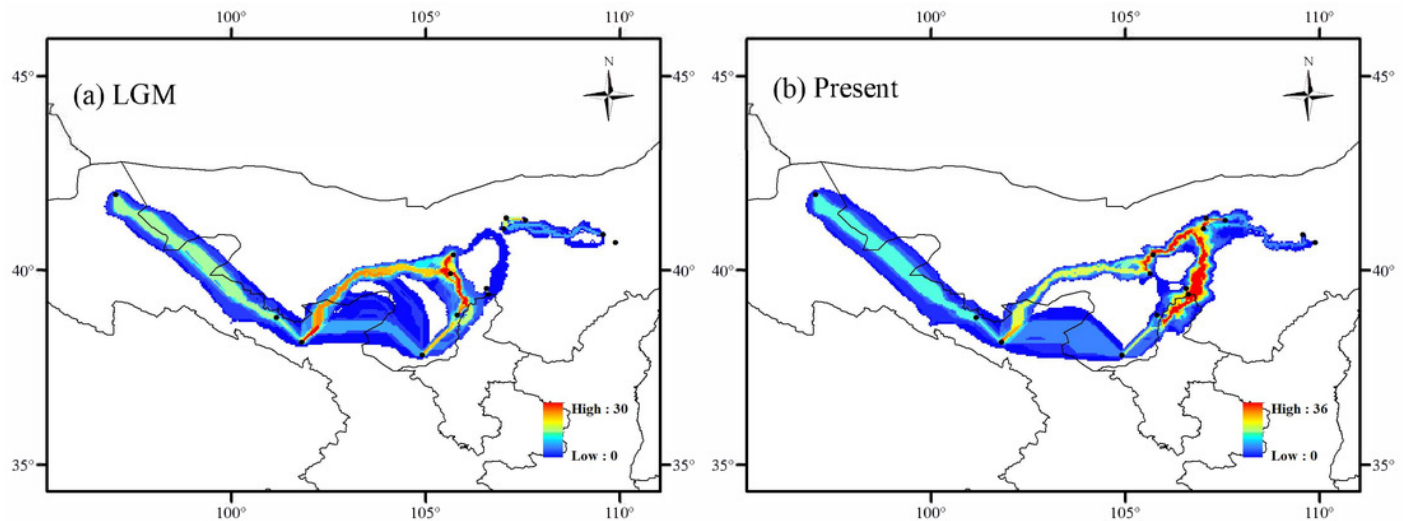
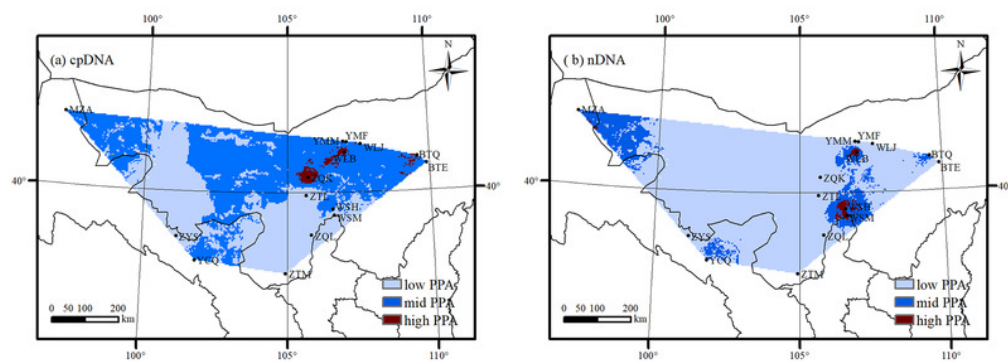



Figure 8

The potential protection area (PPA) of *Amygdalus mongolica* based on genetic diversity (a: cpDNA; b: nDNA).

Sample and genetic variation information for 15 populations of *Amygdalus mongolica*.

1   **Table 1** Sample and genetic variation information for 15 populations of *Amygdalus mongolica*

Population location/ code	Lat / Lon	Sample size	cpDNA haplotype	H_d (± SD)	cpDNA π (± SD*10^{-2})
			nDNA ribotype	R_d (± SD)	nDNA π (± SD*10^{-2})
Western group		**102**		**0.791 ± 0.036**	**0.00683 ± 0.33**
				0.720 ± 0.045	**0.00591 ± 0.21**
MZS	41.95°/97.05°	11	H1 (9), H2 (2)	0.327 ± 0.15	0.00842± 0.39
			R1 (8), R2(3)	0.436 ± 0.13	0.00341 ± 0.21
ZYS	38.78°/101.18°	11	H3 (11)	-	-
			R1 (11)	-	-
YCQ	38.15°/101.83°	11	H3(10), H4 (1)	0.439 ± 0.16	0.00574 ± 0.22
			R1 (8), R3 (3)	0.431 ± 0.01	0.00711 ± 0.30
ZTL	39.92°/105.65°	12	H7 (12)	-	-
			R3 (12)	-	-
ZQK	40.40°/105.72°	12	H5 (7), H6 (1), H7 (4)	0.534 ± 0.06	0.00962± 0.13
			R3 (12)	-	-
ZTM	37.82°/104.93°	12	H8 (12)	-	-
			R3 (12)	-	-
ZQL	38.85°/105.83°	11	H5 (11)	-	-
			R3 (11)	-	-
WSM	39.38°/106.63°	7	H5 (7)	-	-
			R4 (4), R5 (3)	0.509 ± 0.01	0.00683 ± 0.32
WSH	39.54°/106.58°	15	H5 (14), H9 (1)	0.417 ± 0.01	0.00342 ± 0.23
			R4 (9), R5 (6)	0.513 ± 0.008	0.00911 ± 0.39
Eastern group		**72**		**0.804 ± 0.04**	**0.00723 ± 0.24**
				0.576 ± 0.01	**0.00394 ± 0.19**
WLB	41.07°/107.02°	13	H10 (4), H11 (3), H12 (6)	0.567 ± 0.13	0.00544 ± 0.11
			R6 (2), R7 (4), R8 (7)	0.676 ± 0.07	0.00922 ± 0.39
YMM	41.35°/107.08°	13	H12 (13)	-	-
			R8 (13)	-	-
YMF	41.34°/107.09°	12	H11 (4), H12 (8)	0.467 ± 0.132	0.00681 ± 0.32
			R8 (12)	-	-
WLJ	41.29°/107.58°	10	H12 (1), H13 (9)	0.410 ± 0.15	0.00652 ± 0.34
			R8 (10)	-	-
BTQ	40.92°/109.58°	12	H14 (1), H15 (5), H16 (6)	0.530 ± 0.076	0.00682 ± 0.22
			R6 (4), R9 (8)	0.513 ± 0.082	0.00612 ± 0.17
BTE	40.72°/109.90°	12	H15 (5), H17 (7)	0.447 ± 0.032	0.00211 ± 0.13
			R10 (12)	-	-

Estimates of gene diversity and population differentiation (mean ± SD) for the total populations of *Amygdalus mongolica*.

1   **Table 2** Estimates of gene diversity and population differentiation (mean ± SD) for the total populations of *Amygdalus mongolica*

Data set	H_T	H_S	G_{ST}	N_{ST}
cpDNA	0.807 ± 0.050	0.225 ± 0.053	0.712 ± 0.070	0.858 ± 0.083
nDNA	0.815 ± 0.077	0.212 ± 0.014	0.690 ± 0.036	0.829 ± 0.013

2   H_T, total gene diversity; H_S, average gene diversity within populations; G_{ST} and N_{ST}, population differentiation values

Analysis of molecular variance (AMOVA) for 15 populations of *Amygdalus mongolica.*

1 **Table 3** Analysis of molecular variance (AMOVA) for 15 populations of *Amygdalus mongolica*

Data set	Source of variation	d. f.	Sum of squares	Variance components	Percentage of variation (%)	Fixation index
cpDNA	Among populations	14	612.031	4.356	83.87	
	Within populations	159	114.769	0.838	16.13	$F_{CT} = 0.839$
	Total	173	726.800	5.194		
	Among groups	5	421.824	2.862	66.25	$F_{SC} = 0.731$
	Among populations within groups	10	74.343	1.066	24.67	$F_{ST} = 0.91$
	Within populations	159	53.376	0.392	0.393	$F_{CT} = 0.663$
	Total	173	549.544	4.320		
nDNA	Among populations	14	163.713	1.166	79.33	
	Within populations	159	41.334	0.034	20.67	$F_{CT} = 0.793$
	Total	173	205.407	1.470		
	Among groups	5	149.993	1.116	73.14	$F_{SC} = 0.314$
	Among populations within groups	10	10.018	0.12	8.44	$F_{ST} = 0.816$
	Within populations	159	38.257	0.281	18.43	$F_{CT} = 0.731$
	Total	173	198.268	1.527		

2 F_{ST}, correlation within populations relative to the total; F_{SC}, correlation within populations relative to groups; F_{CT}, correlation of haplotypes/

3 ribotypes within groups relative to the total

Results of neutrality tests and mismatch distribution analysis for the overall populations and two regional groups based on the cpDNA and nDNA dataset.

1   **Table 4** Results of neutrality tests and mismatch distribution analysis for the overall populations and two regional groups based on

2   the cpDNA and nDNA dataset

Group	cpDNA dataset						nDNA dataset					
	SSD	P	H_{rag}	P	Fu's Fs	Tajima's D	SSD	P	H_{rag}	P	Fu's Fs	Tajima's D
Overall	0.184	0.072	0.337	0.391	2.365	1.3045	0.186	0.054	0.371	0.317	1.189	1.240
Eastern group	0.225	0.063	0.3	0.457	0.665	2.6346	0.197	0.245	0.351	0.354	2.734	1.154
Western group	0.091	0.126	0.335	0.303	6.495	-0.5173	0.191	0.079	0.397	0.297	1.142	1.020

3   *SSD*, sum of squared deviations; H_{Rag}, Harpending's raggedness index.