



OPEN

## Phylogeography of *Prunus armeniaca* L. revealed by chloroplast DNA and nuclear ribosomal sequences

Wen-Wen Li<sup>1</sup>, Li-Qiang Liu<sup>1</sup>, Qiu-Ping Zhang<sup>2</sup>, Wei-Quan Zhou<sup>1</sup>, Guo-Quan Fan<sup>3</sup> & Kang Liao<sup>1</sup>✉

To clarify the phylogeography of *Prunus armeniaca* L., two chloroplast DNA fragments (*trnL-trnF* and *ycf1*) and the nuclear ribosomal DNA internal transcribed spacer (ITS) were employed to assess genetic variation across 12 *P. armeniaca* populations. The results of cpDNA and ITS sequence data analysis showed a high level of genetic diversity (cpDNA:  $H_T = 0.499$ ; ITS:  $H_T = 0.876$ ) and a low level of genetic differentiation (cpDNA:  $F_{ST} = 0.1628$ ; ITS:  $F_{ST} = 0.0297$ ) in *P. armeniaca*. Analysis of molecular variance (AMOVA) revealed that most of the genetic variation in *P. armeniaca* occurred among individuals within populations. The value of interpopulation differentiation ( $N_{ST}$ ) was significantly higher than the number of substitution types ( $G_{ST}$ ), indicating genealogical structure in *P. armeniaca*. *P. armeniaca* shared genotypes with related species and may be associated with them through continuous and extensive gene flow. The haplotypes/genotypes of cultivated apricot populations in Xinjiang, North China, and foreign apricot populations were mixed with large numbers of haplotypes/genotypes of wild apricot populations from the Ili River Valley. The wild apricot populations in the Ili River Valley contained the ancestral haplotypes/genotypes with the highest genetic diversity and were located in an area considered a potential glacial refugium for *P. armeniaca*. Since population expansion occurred 16.53 kyr ago, the area has provided a suitable climate for the population and protected the genetic diversity of *P. armeniaca*.

The evolutionary history of organisms, including their genetic diversity, population structure and historical dynamics, is critical to species conservation<sup>1</sup>. Understanding the effects of climate change on the spatial genetic patterns of species, particularly endangered species, can help reveal not only the evolutionary history of species but also conservation strategies<sup>2,3</sup>. The origins of mountain biodiversity are complex and may include the immigration of preadapted lineages<sup>4–5</sup>, in situ diversification<sup>6</sup>, or the continuation of ancestral lineages<sup>7</sup>. Compared with those of other mountains, such as the Hengduan Mountains, the organismal evolution and diversity of the Tianshan Mountains are still poorly understood. The Ili River Valley is located in the western part of the Tianshan Mountains in China and is surrounded by mountains on three sides. This valley was the main northern crossing of the ancient Silk Road. In the late Tertiary period, a large number of species, including wild apricot, wild apple and wild hawthorn, remained in the Ili River Valley and were important components of the deciduous broad-leaved forest at elevations below the coniferous forest and above the mountain grassland belt in the Xinjiang Uygur Autonomous Region, China<sup>8</sup>. However, there have been few studies on the phylogeography of plant species in the arid region of Northwest China<sup>9–10</sup>.

In recent decades, the method of combining molecular data with paleoclimatic and geographical evidence has been effective in the study of systematic geography<sup>11–12</sup>. Many systematic geographic studies have shown that the Last Glacial Maximum (LGM) of the Pleistocene strongly influenced the genetic variation and biodiversity of plants throughout the Northern Hemisphere<sup>13–14</sup>. Especially during the Quaternary glacial period, species in the ice-free zone were mainly affected by the cold and dry climate<sup>15</sup>. Climatic fluctuations cause the distribution of species to shrink and expand<sup>16</sup>, and cold and dry climates prompt plant and animal species to retreat to refugia,

<sup>1</sup>College of Horticulture and Forestry, Xinjiang Agricultural University, Urumqi, Xinjiang, China. <sup>2</sup>Xiongyue National Germplasm Resources Garden of the Liaoning Institute of Pomology, Xiongyue, Shenyang, China. <sup>3</sup>Luntai National Fruit Germplasm Resources Garden of Xinjiang Academy of Agricultural Sciences, Luntai, Xinjiang, China. ✉email: liaokang01@163.com

which provided shelter<sup>17</sup>. As temperatures rose after the glacial period, species underwent population expansion, changes that left a genetic signature in their current populations<sup>18</sup>. Although no unified glacial sheet was present in Asia, paleodata suggest that the distribution ranges of woody species in this region were similarly changed by Quaternary climatic oscillations<sup>19</sup>. Based on the geographical distribution of genetic variation, the glacial refuges and postglacial revegetation routes of most woody plants are roughly consistent with fossil evidence<sup>20</sup>. However, due to the lack of fossil evidence, genetic evidence has become an important means of providing information on the distribution range of some woody species and the history of glacial refugia<sup>21–22</sup>. Volkova<sup>23</sup> interpreted the observed genetic structure [nuclear (internal transcribed spacer, ITS) and plastid DNA] of the Eurasian populations of *Prunus padus* as plausibly resulting from at least two cycles of glacial survival in refugia followed by postglacial colonization events<sup>17</sup>. The complex geographic history in the area may have provided a refuge for species in the last glacial period<sup>17</sup>. Xu et al.<sup>24</sup> proposed two possible independent glacial refugia in Northwest China: the Ili River Valley and the Northern Junggar Basin. Su et al.<sup>9</sup> speculated that intensification of the dry and cold climate during the early Quaternary, combined with plant physiological features, contributed to the lineage split, and climate oscillations most likely led to the Ili range expansion. Therefore, it is of great interest to study the systematic geography of endemic plants in Northwest China against the background of Quaternary climatic fluctuations.

Apricot is a fruit of temperate and subtropical regions. Turkey, Uzbekistan, Iran, Algeria, Italy, Spain, China, Pakistan, France and Japan are the main producers of apricots (<http://www.fao.org/home/zh>). Apricot belongs to section *Armeniaca* (Lam.) Koch, subgenus *Prunophora* Focke and genus *Prunus* (Rosaceae)<sup>25</sup>. Almost all cultivated apricots originated from *Prunus armeniaca*<sup>26</sup>. In China, wild apricot is distributed in the Ili River Valley (Tianshan Mountain area). The species is also distributed along the Tianshan Mountains and westward to Kazakhstan, Kyrgyzstan and Uzbekistan. It is a relic of a broad-leaved forest from the late Tertiary period, which played a decisive role in the domestication of cultivated apricots worldwide<sup>27</sup>. As the world experienced extreme weather events, especially glacial periods, and most species went extinct, some of the more complex valleys may have become refugia for surviving forests and local species. As a result, traces of the species may be gradually shrinking in such areas. Therefore, is the Ili River Valley a glacial refugium for wild apricot?

Generally, the dispersal distance of seeds is much shorter than that of pollen, and population divergence due to genetic drift will be more marked for chloroplast DNA (cpDNA) than for nuclear DNA. Indeed, cpDNA is considered to evolve very slowly, with low recombination and mutation rates<sup>28</sup>. Organelle markers could provide powerful tools for studying the phylogeography and migratory footprints of species<sup>29</sup>. Parental genetic markers are often combined with single-parent organelle markers for population genetics studies<sup>30</sup>. cpDNA and ITS sequence variations have been very effective in revealing the glacial refuges of plants<sup>29</sup>. cpDNA lineages usually show the unique geographical distribution and evolutionary history of natural populations and are therefore widely used in systematic geography studies<sup>31</sup>. Li et al.<sup>32</sup> successfully used cpDNA and ITS markers to evaluate the diversity and phylogenetic relationships of populations of *Saxifraga sinomontana*, indicating that the species had microrefugia during the Quaternary glacial period. In this study, we employed cpDNA (*trnL-trnF* and *ycf1*) and nuclear ribosomal DNA (nrDNA) sequences to (1) reveal the haplotype/genotypic diversity and population genetic structure of the species and (2) examine the demographic history of *P. armeniaca* during Quaternary climate oscillations, and further explore the origin and evolution of this species.

## Materials and methods

**Sample collection.** The samples used for cpDNA analysis included 123 individuals from 20 populations. A total of 171 individuals from 19 populations were used for the ITS analysis, of which 38 samples were obtained from the NCBI database (Table S1). The samples studied were from *P. armeniaca* and related species (*Prunus sibirica*, NAG; *Prunus mandshurica*, LX; *Prunus dasycarpa*, ZX; *Prunus mume*, ECG; *Prunus zhengheensis*, ZHX<sup>33</sup>; *Prunus limeixing*, LMX<sup>33</sup> and *Prunus brigantina*, FGX). *Prunus davidiana* (T) was used as the outgroup.

Our collection of wild apricot (*P. armeniaca*) populations covered most of the natural distribution in China, including Huocheng County (DZGhcmd, DZGhcy and DZGhcm populations), Yining County (DZGyn population), Gongliu County (DZGglb and DZGgld populations) and Xinyuan County (DZGxyt, DZGxya and DZGxyz populations). The distance between individuals sampled in each population was at least 100 m. Young leaves were collected and dried immediately with silica gel.

The cultivated populations of *P. armeniaca* included the Xinjiang apricot group (CAG, cultivated apricots in Xinjiang), the North China apricot group (NCG, cultivated apricots in Shandong, Shaanxi, Gansu, Liaoning and Ningxia) and the foreign apricot group (EG, cultivated apricots in the USA, France, Italy and Australia). Detailed sample information is provided in Table 1 and Table S2. The main characteristics of different populations can be found in Zhang et al.<sup>8</sup>.

**DNA sequencing.** Total genomic DNA was extracted from the silica gel-dried leaf materials using a Plant Genomic DNA Kit (Tiangen Biotech, Beijing, China)<sup>34</sup>. The quality and concentration of the extracted DNA were determined by 1% agarose gel electrophoresis and ultraviolet spectrophotometry, respectively.

cpDNA and nrDNA sequences from 15 samples were initially screened using universal primers. The sequencing results showed that the sequences of cpDNA (genes *trnL-trnF* and *ycf1*) and two nuclear ribosomal ITS regions (ITS1 and ITS2) were polymorphic. cpDNA and ITS fragments were amplified by polymerase chain reaction (PCR), and the details of their primers are provided in Table S3<sup>35–37</sup>. PCR was performed in a total volume of 25  $\mu$ L that contained 1  $\mu$ L DNA, 5.5  $\mu$ L PCR mix, 16.5  $\mu$ L double-distilled water and 1  $\mu$ L each forward or reverse primer. PCR amplifications were performed under the following conditions: 5 min of initial denaturation at 94 °C and 35 cycles of 0.5 min at 94 °C, 0.5 min of annealing at 58°, and 0.5 min of extension at 72 °C, with 10 min of final extension at 72 °C. A CASpure PCR Purification Kit (CASarray, Shanghai, China)<sup>32</sup> was used for

Population	Origin/location	Type	Latitude (N°)	Longitude (E°)	Altitude (m)
<b><i>P. armeniaca</i></b>					
DZGhcmd	Huocheng, Xinjiang	W	44.43	80.79	1187.6
DZGhcy	Huocheng, Xinjiang	W	44.44	80.79	1245.7
DZGhcm	Huocheng, Xinjiang	W	44.40	80.71	1244.1
DZGyn	Yining, Xinjiang	W	44.12	81.62	1983.6
DZGgib	Gongliu, Xinjiang	W	43.25	82.86	1371.5
DZGgld	Gongliu, Xinjiang	W	43.23	82.75	1269.4
DZGxyt	Xinyuan, Xinjiang	W	43.54	83.44	1138.8
DZGxya	Xinyuan, Xinjiang	W	43.50	83.70	1275.3
DZGxyz	Xinyuan, Xinjiang	W	43.38	83.61	1374.1
CAG	Luntai, Xinjiang	C	41.78	84.23	972.0
NCG	Luntai, Xinjiang	C	41.78	84.23	972.0
EG	Xiongyue, Liaoning	C	40.17	122.16	20.4
<b><i>P. sibirica</i></b>					
NAG	Xiongyue, Liaoning	W	40.17	122.16	20.4
<b><i>P. mandshurica</i></b>					
LX	Xiongyue, Liaoning	C	40.17	122.16	20.4
<b><i>P. dasycarpa</i></b>					
ZX	Luntai, Xinjiang	C	41.78	84.23	972.0
<b><i>P. mume</i></b>					
ECG	Xiongyue, Liaoning	C	40.17	122.16	20.4
<b><i>P. zhengheensis</i></b>					
ZHX	Xiongyue, Liaoning	C	40.17	122.16	20.4
<b><i>P. limeixing</i></b>					
LMX	Xiongyue, Liaoning	C	40.17	122.16	20.4
<b><i>P. brigantina</i></b>					
FGX	Xiongyue, Liaoning	C	40.17	122.16	20.4
<b><i>P. davidiana</i></b>					
T	Luntai, Xinjiang	O	41.78	84.23	972.0

**Table 1.** Population code, sampling location, coordinates, altitude and types of *P. armeniaca* and relative species. W, wild; C, cultivars; O, out group.

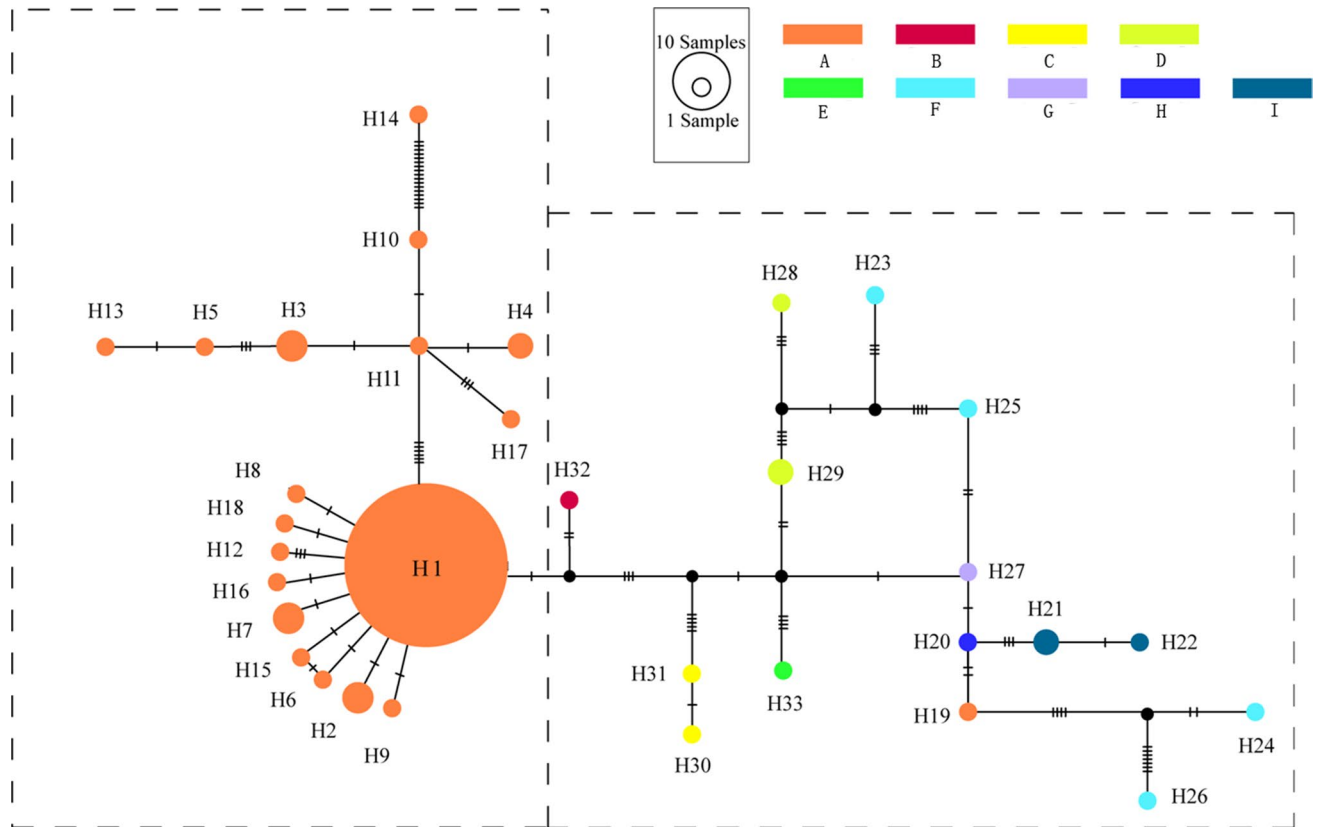
purification. The purified PCR products were sequenced on an ABI PRISM 3730XL DNA Analyzer (Applied Biosystems, Foster City, CA, USA)<sup>38</sup>.

**Data analysis.** We used BioEdit<sup>39</sup> to view and manually correct the sequencing results. We first used CLUSTAL W<sup>40</sup> to align the sequences and coded indels following the method of Simmons and Ochoterena<sup>41</sup>. Then, manual adjustments were made in MEGA ver. 7.0.26<sup>42</sup> to remove the overhanging tails and ensure a uniform sequence length. We concatenated two chloroplast fragments (*trnL-trnF* and *ycf1*) into a separate matrix for subsequent analysis. We used DnaSP ver. 5.10 to identify different haplotypes (cpDNA sequences) or genotypes (ITS sequences). A haplotype network was constructed using TCS ver. 1.2.1<sup>43</sup>. ArcGIS ver. 10.2 (<http://desktop.arcgis.com>) software was used to create a haplotype geographical distribution map.

Haplotype/genotype diversity ( $H_d$ ) and nucleotide diversity ( $\pi$ ) were calculated using DnaSP ver. 5.10 software<sup>44</sup>. The within-population gene diversity ( $H_s$ ), gene diversity in all populations ( $H_T$ ), interpopulation differentiation ( $G_{ST}$ ) and number of substitution types ( $N_{ST}$ ) were calculated using PERMUT<sup>45</sup> ver. 2.0. The last two indexes ( $G_{ST}$  and  $N_{ST}$ ) were analyzed via permutation tests with 1000 permutations. When  $N_{ST}$  is greater than  $G_{ST}$ , it indicates the existence of genealogical geographic structure<sup>45</sup>. Analysis of molecular variance (AMOVA) was performed using Arlequin ver. 3.5.2.2<sup>46</sup> to partition the genetic variation at different levels, with statistical significance determined by 1,000 permutations.

To investigate the historical dynamics of *P. armeniaca*, mismatch distribution analysis was conducted using DnaSP ver. 5.10. The sum of squared deviations (SSDs), Harpending's raggedness index (HRI)<sup>45</sup> and corresponding  $P$  values were calculated using Arlequin ver. 3.5.2.2<sup>46</sup>. Neutrality tests based on Tajima's  $D$  and Fu's  $F_S$  were conducted to detect departures from the population equilibrium by Arlequin ver. 3.5.2.2<sup>46</sup>. According to the formula of Rogers and Harpending<sup>47</sup>,  $T = \tau/2u$ , where " $\tau$ " is the parameter value from the mismatch distribution model. In the formula  $u = \mu kg$ , " $\mu$ " is the base substitution rate (chloroplast angiosperms<sup>48</sup>:  $1.1 \times 10^{-9}$ ), " $k$ " is the fragment length (cpDNA length after combination: 2062 bp), and " $g$ " is the generation time (20 years)<sup>49</sup>.

The outgroup was *P. davidiana*, and the time point of peach-apricot differentiation was used as the calibration point<sup>50</sup>. The BEAUti interface was used to create an input file for BEAST<sup>51</sup>, for which the GTR + I + G nucleotide substitution model was used. The data were analyzed using a relaxed log-normal clock model and the Yule



**Figure 1.** The haplotype network generated from the haplotypes of *P. armeniaca* and related species based on cpDNA dataset. The small black circles shown an intermediate haplotype not detected in this study. (A) *P. armeniaca*; (B) *P. zhengheensis*; (C) *P. davidiana*; (D) *P. sibirica*; (E) *P. dasycarpa*; (F) *P. limeixing*; (G) *P. mandshurica*; (H) *P. mume*; (I) *P. brigantina*. The haplotype network was constructed using TCS ver. 1.2.1 and then improved in Adobe Illustrator CC (Adobe Systems Inc., CA, USA).

process speciation model for the tree priors. Prior settings for calibrating nodes were an offset of 55.1 Ma and a log mean of 1.0 (log stdev of 0.5). The Bayesian Markov chain Monte Carlo simulation was run for 100 million generations with a sample frequency of 1000, and the first 20% of generations were discarded as burn-in. Three independent analyses were conducted and their results combined by LogCombiner ver. 1.8.4. Finally, annotation and visualization of the maximum clade credibility tree were performed in TreeAnnotator 1.8.4 and FigTree ver. 1.4.3, respectively.

## Results

**Haplotype/genotype phylogenetics and distribution.** Based on the concatenated cpDNA sequences (*trnL-trnF* and *ycf1*), 33 haplotypes (H1–H33) were identified among 123 individuals from 20 populations of *P. armeniaca* and related species (Fig. 1, Table 2). The alignment lengths of the two chloroplast fragments were 746 bp and 1316 bp, respectively, and the combined length was 2026 bp. Variable sites among the 33 haplotypes are shown in Table S4. *P. armeniaca* harbored haplotypes H1–H17 and H19–H20; *P. sibirica* harbored haplotypes H28–H29; *P. mandshurica* harbored haplotypes H27; *P. dasycarpa* harbored haplotypes H33; *P. mume* harbored haplotypes H20; *P. zhengheensis* harbored haplotypes H32; *P. limeixing* harbored haplotypes H23–H26; *P. brigantina* harbored haplotypes H21–H22; and *P. davidiana* harbored haplotypes H30–H31. The results showed that populations of different species did not share haplotypes. The results showed that, except for the EG accessions, the diversity of wild *P. armeniaca* (DZG accessions) (nucleotide diversity and haplotype diversity: 0.0013, 0.444, respectively) was higher than that of the CAG (0.0006, 0.239) and NCG (0.0002, 0.400) accessions.

In the ITS dataset, a total of 57 ITS genotypes were discovered among 171 individuals from 19 populations of *P. armeniaca* and its related species (Table 2, Figure S1), and the alignment length was 545 bp. The variable sites among the 57 genotypes are shown in Table S5. *P. armeniaca* harbored haplotypes T1–T30, T42 and T48–T50; *P. sibirica* harbored haplotypes T2–T3, T8–T9, T12, T25 and T44–T47; *P. mandshurica* harbored haplotypes T3 and T9; *P. dasycarpa* harbored haplotypes T2 and T55–T57; *P. mume* harbored haplotypes T11 and T31–T41; *P. zhengheensis* harbored haplotypes T52–T54; *P. limeixing* harbored haplotypes T3 and T43; and *P. davidiana* harbored haplotypes T51. Except for the EG accessions, the diversity of wild *P. armeniaca* (DZG accessions) (nucleotide diversity and haplotype diversity: 0.0055 and 0.878, respectively) was higher than that of the CAG (0.0053, 0.832) and NCG (0.0047 and 0.818) accessions (Fig. 2).

A phylogenetic tree of all 57 genotypes was constructed to better understand their relationships (Fig. 3). The phylogenetic tree roughly divided the collected accessions into two groups. One group included the *P. armeniaca*

Population	cpDNA				ITS			
	Haplotype composition	N	Hd	$\pi$	Genotype composition	N	Hd	$\pi$
DZGcmd	H1(6), H10(1), H11(1), H12(1)	9	0.583	0.0016	T2(2), T9(1), T14(1), T17(2), T19(1), T20(1), T21(1), T22(1)	10	0.956	0.0082
DZGhcy	H1(9)	9	0.000	0.0000	T3(2), T8(1), T9(3), T23(1),	7	0.810	0.0067
DZGhcm	H1(7), H8(1), H9(1)	9	0.417	0.0002	T2(1), T3(4), T8(1), T9(1), T18(1)	8	0.786	0.0000
DZGyn	H1(6), H7(1), H15(1), H16(1), H17(1)	10	0.667	0.0011	T1(1), T3(2), T8(2), T9(1), T14(2), T30(1)	9	0.917	0.0087
DZGglb	H1(6), H5(1), H6(1), H7(2)	10	0.644	0.0012	T2(1), T3(3), T8(2), T13(1), T14(1), T15(1)	9	0.889	0.0043
DZGgld	H1(6), H3(1)	7	0.286	0.0010	T3(4), T8(1), T16(1), T17(1)	7	0.714	0.0027
DZGxyt	H1(8), H2(1)	9	0.222	0.0001	T3(4), T8(2), T9(2), T12(1), T25(1)	10	0.822	0.0032
DZGxya	H1(1), H4(1), H13(1)	3	1.000	0.0039	T2(2), T24(1)	3	0.667	0.0988
DZGxyz	H1(7), H3(1), H14(1)	9	0.417	0.0031	T2(2), T3(2), T26(1), T27(1), T28(1), T29(2)	9	0.917	0.0057
Within population		75	0.444	0.0013		72	0.878	0.0055
CAG	H1(21), H2(1), H3(1), H4(1)	24	0.239	0.0006	T1(5), T2(11), T3(8), T4(1), T5(1), T6(1), T7(1), T8(1), T9(1), T10(1), T11(2), T12(1)	34	0.832	0.0053
NCG	H1(4), H2(1)	5	0.400	0.0002	T1(1), T3(5), T9(2), T48(1), T49(1), T50(2)	12	0.818	0.0047
EG	H1(1), H18(1), H19(1)	3	1.000	0.0045	T2(1), T3(4), T42(1)	6	0.600	0.0053
Within population		32	0.343	0.0012		52	0.833	0.0052
Among population		107	0.548	0.0019		124	0.865	0.0054
NAG	H28(1), H29(2)	3	0.667	0.0023	T2(1), T3(6), T8(1), T9(1), T12(1), T25(1), T44(1), T45(1), T46(1), T47(1)	15	0.857	0.0104
LX	H27(1)	1	0.000	0.0000	T3(4), T9(1)	5	0.400	0.0014
ZX	H33(1)	1	0.000	0.0000	T2(2), T55(1), T56(2), T57(2)	7	0.857	0.0229
ECG	H20(1)	1	0.000	0.0000	T11(1), T31(1), T32(1), T33(1), T34(1), T35(1), T36(1), T37(1), T38(1), T39(1), T40(1), T41(1)	12	1.000	0.017
ZHX	H32(1)	1	0.000	0.0000	T52(1), T53(1), T54(1)	3	1.000	0.0047
LMX	H23(1), H24(1), H25(1), H26(1)	4	1.000	0.0051	T3(3), T43(1)	4	0.500	0.0170
FGX	H21(2), H22(1)	3	0.667	0.0003				
T	H30(1), H31(1)	2	1.000	0.0005	T51(1)	1	0.000	0.0000

**Table 2.** Sample information and summary of haplotype/genotype distribution, genetic diversity for each population. N, sample size; Hd, haplotype/genotype diversity;  $\pi$ , nucleotide diversity.

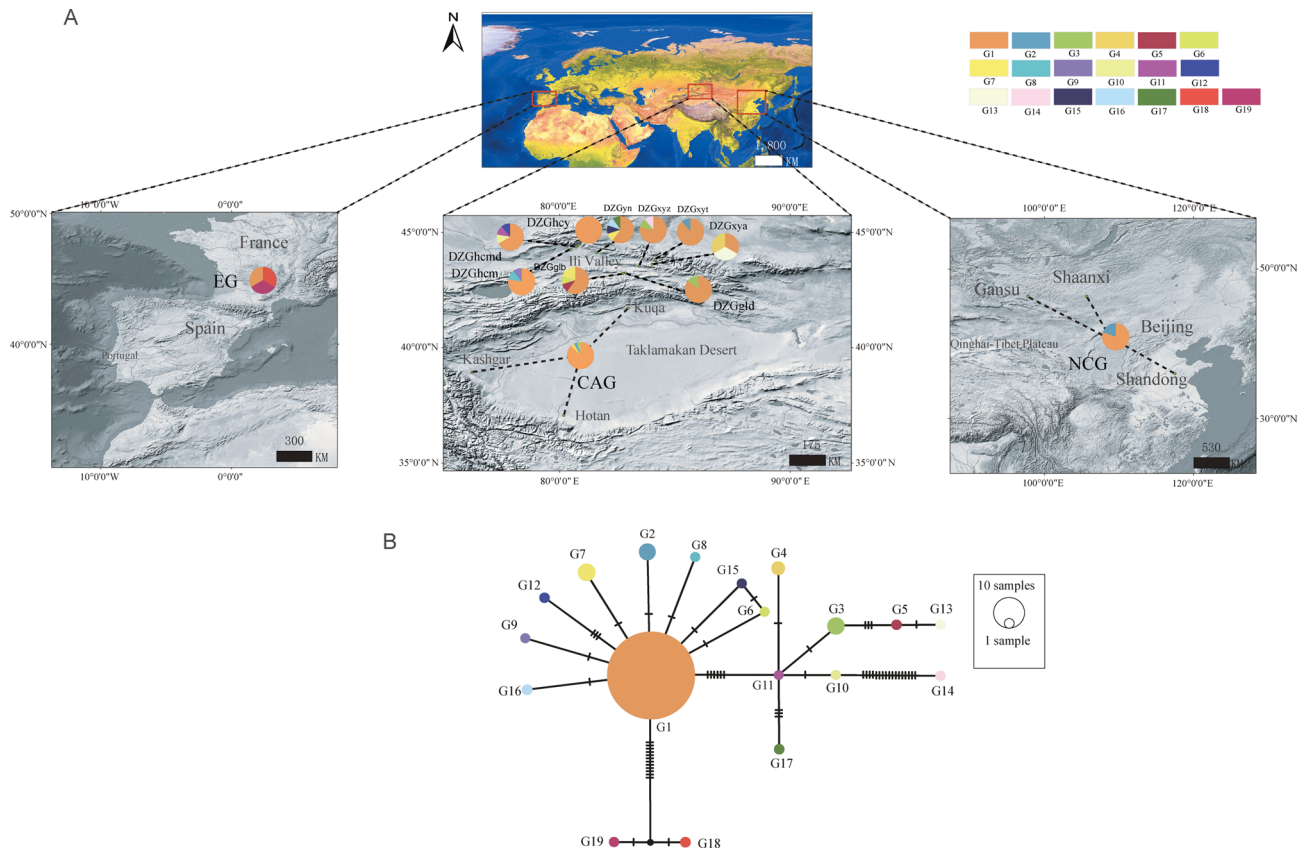
(blue branch); the second group was composed of related species (green branch), including *P. sibirica*, *P. mandshurica*, *P. dasycarpa*, *P. mume*, *P. zhengheensis* and *P. limeixing*. T3 was widely distributed in most populations. The genetic backgrounds of the related species had the same genotypes (T2, T3, T11, T25, T8, T9, and T12) as *P. armeniaca*, indicating that they are associated with *P. armeniaca* through continuous and extensive gene flow.

Based on the concatenated cpDNA sequences (*trnL-trnF* and *ycf1*), 19 haplotypes (G1-G19) were identified among 107 individuals from 12 populations of *P. armeniaca* (Fig. 2). Variable sites among the 19 haplotypes are shown in Table S6. The Hd and  $\pi$  values detected at the cpDNA sequence level in *P. armeniaca* were 0.548 and  $1.9 \times 10^{-3}$ , respectively. The geographic distribution of the 19 haplotypes is shown in Fig. 2A, illustrating that G1 haplotypes were distributed in all populations. The cpDNA haplotype network (Fig. 2B) showed that 7 haplotypes were differentiated from haplotype G1 by a one-step mutation with G1 at the center. Three haplotypes were differentiated from G11 by a one-step mutation. The overall network map showed a "star-like" distribution pattern. Based on the ITS sequences, 34 haplotypes (P1-P34) were identified among 124 individuals from 12 populations of *P. armeniaca* (Figure S2). The Hd and  $\pi$  values detected at the ITS sequence level in *P. armeniaca* were 0.865 and  $5.4 \times 10^{-3}$ , respectively.

**Population structure and genealogical geography.** In *P. armeniaca*, the gene diversity among populations (cpDNA:  $H_T=0.499$ ; ITS:  $H_T=0.876$ ) was higher than that within populations (cpDNA:  $H_S=0.490$ ; ITS:  $H_S=0.794$ ) (Table 3). A permutation test showed that  $N_{ST}$  was significantly higher than  $G_{ST}$  (cpDNA:  $N_{ST}=0.227 > G_{ST}=0.020$ ; ITS:  $N_{ST}=0.126 > G_{ST}=0.094$ ;  $P < 0.05$ ), indicating that *P. armeniaca* has significant geographical structure (Table 3).

AMOVA revealed significant genetic differentiation among all populations of *P. armeniaca* (cpDNA:  $F_{ST}=0.1628$ ,  $P < 0.001$ ; ITS:  $F_{ST}=0.0297$ ,  $P < 0.001$ ), with most of the genetic diversity occurring within the populations and relatively little occurring among them (Table 4).

**Demographic history and estimation of divergence times.** The mismatch distribution analysis based on the cpDNA and ITS dataset analysis, in which multimodal data were drawn from the cultivated populations or all populations, revealed a demographic equilibrium (Fig. 4, Figure S3). Both the neutrality tests based on Tajima's *D* (cpDNA:  $-2.272$ ,  $P < 0.05$ ; ITS:  $-0.966$ ,  $P < 0.05$ ) and Fu's  $F_S$  (cpDNA:  $-5.8253$ ,  $P < 0.05$ ; ITS:  $-2.223$ ,  $P < 0.05$ ) and the mismatch distribution analysis (Fig. 4) based on the cpDNA and ITS datasets suggested recent range or demographic expansion in wild populations of *P. armeniaca*. In addition, neither the SSDs (cpDNA:  $0.037$ ,  $P > 0.05$ ; ITS:  $0.002$ ,  $P > 0.05$ ) nor the HRI (cpDNA:  $0.179$ ,  $P > 0.05$ ; ITS:  $0.017$ ,  $P > 0.05$ )



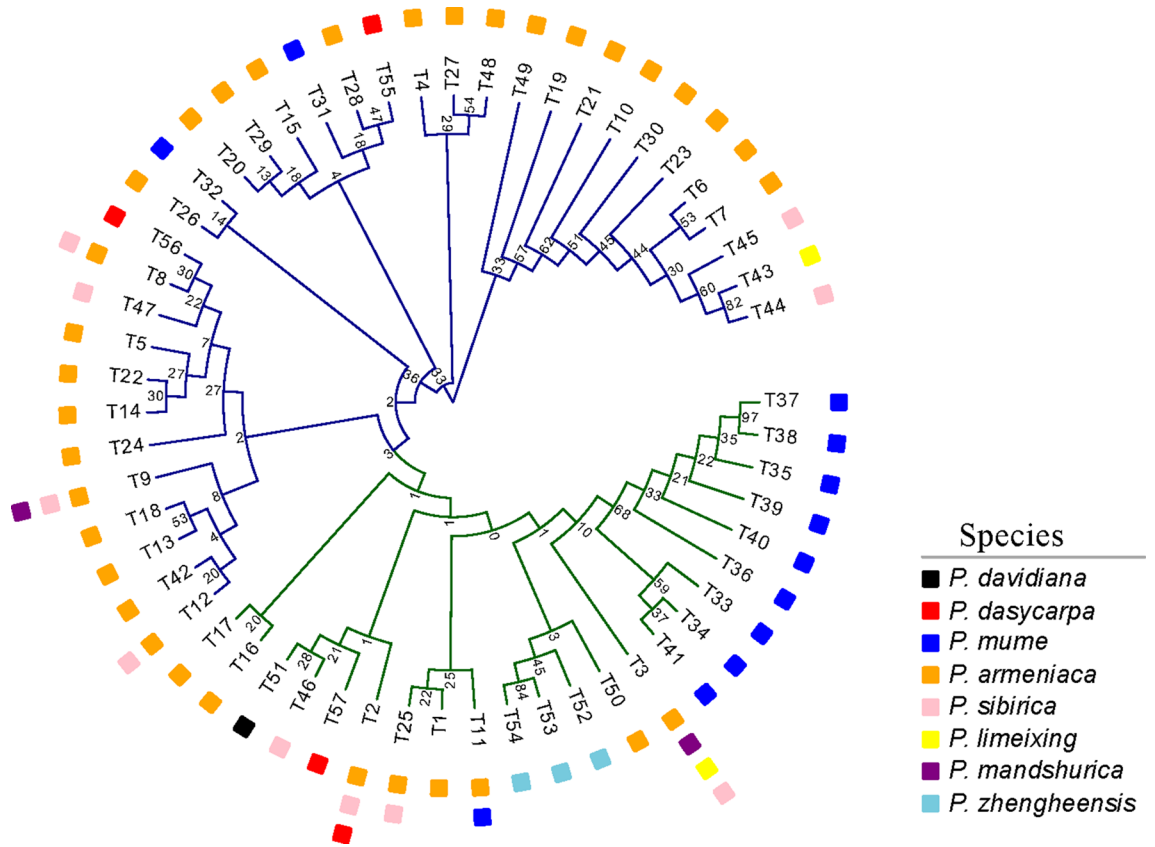
**Figure 2.** Geographic location and haplogroup distribution patterns of the 12 populations of *P. armeniaca* based on cpDNA dataset. (A) Geographic distribution of haplotypes of *P. armeniaca*; Pie chart size corresponds to the sample size of each population. (B) The haplotype network generated from the haplotypes of *P. armeniaca*. The haplotype network generated from the 19 haplotypes of *P. armeniaca*. The small black circles shown an intermediate haplotype not detected in this study. All maps (Source: <http://ditu.ps123.net/world/2363.html>) were generated using ArcGIS ver. 10.V.10.2.2 (ESRI, CA, USA) and then improved in Adobe Illustrator CC (Adobe Systems Inc., CA, USA).

showed no significant positive values (Table 5), indicating no deviation of the observed mismatch distribution from that obtained via model simulation under sudden demographic expansion. Thus, we concluded that the demographic expansion of the wild populations of *P. armeniaca* occurred 16.53 kyr ago.

The cpDNA dataset was employed to estimate when the onset of divergence between *P. armeniaca* and its related species occurred (Fig. 5, Table S7). Thirty-three haplotypes were divided into two groups: those of *P. armeniaca* (blue) and those of related species (green) (Fig. 5). The divergence time estimation revealed that the differentiation of *P. armeniaca* from its related species occurred during the middle Eocene, approximately 45.68 Ma (95% highest posterior density (HPD) = 28.47–61.87). The onset of intraspecific divergence in *P. armeniaca* was estimated to have occurred 25.55 (95% HPD = 12.93–39.63) Ma.

## Discussion

Parental genetic markers are often combined with single-parent organelle markers for population genetics studies. Li et al.<sup>32</sup> used cpDNA *trnL-trnF*, *rpl16* and nrDNA ITS sequences to infer the evolutionary history of *S. sinomontana*. Yang et al.<sup>52</sup> used cpDNA *psbA-trnH*, *trnL-trnE*, *ycf1*, and *matK* sequences to access the demographical history and genetic diversity of a Deciduous Oak (*Quercus liaotungensis*) in Northern China. Zhang et al.<sup>53</sup> used cpDNA markers to successfully determine the genetic diversity, genetic structure, and demographic history of 7 *Michelia yunnanensis* populations. Many scholars<sup>54–57</sup> believed that the diversity of wild apricot is richest in the Ili River Valley, with low levels of genetic differentiation and genetic variation mainly occurring within populations. Hu et al.<sup>56</sup> used simple sequence repeat markers to analyze the diversity of 212 apricot germplasms from 14 populations in the Ili River Valley. Among the populations, that from the Tuergen township in Xinyuan County had the highest genetic diversity, and the genetic distance between populations was significantly correlated with geographical distance. The self-incompatibility, wide distribution, and long-distance transmission of pollen through insects and strong winds of apricot are the main factors affecting its genetic structure.<sup>56</sup> Based on cpDNA and ITS data, we concluded that the haplotype/genotype diversity of wild apricot populations distributed in Ili River Valley was relatively high (Table 2), with that of the DZGhcmd and DZGyn populations being the highest. The results of AMOVA (Table 4) showed that the genetic diversity in *P. armeniaca* mainly occurs within populations (cpDNA: 83.72%; ITS: 97.03%), but there were also significant differences among populations (cpDNA:



**Figure 3.** The Neighbor-joining consensus tree (NJ) of *Prunus* spp. based on genotypes of ITS sequences. The numbers on branches indicate the bootstraps values.

Populations	cpDNA				ITS			
	$H_S$	$H_T$	$G_{ST}$	$N_{ST}$	$H_S$	$H_T$	$G_{ST}$	$N_{ST}$
<i>P. armeniaca</i> (cultivated & wild)	0.490	0.499	0.020	0.227*	0.794	0.876	0.094	0.126*

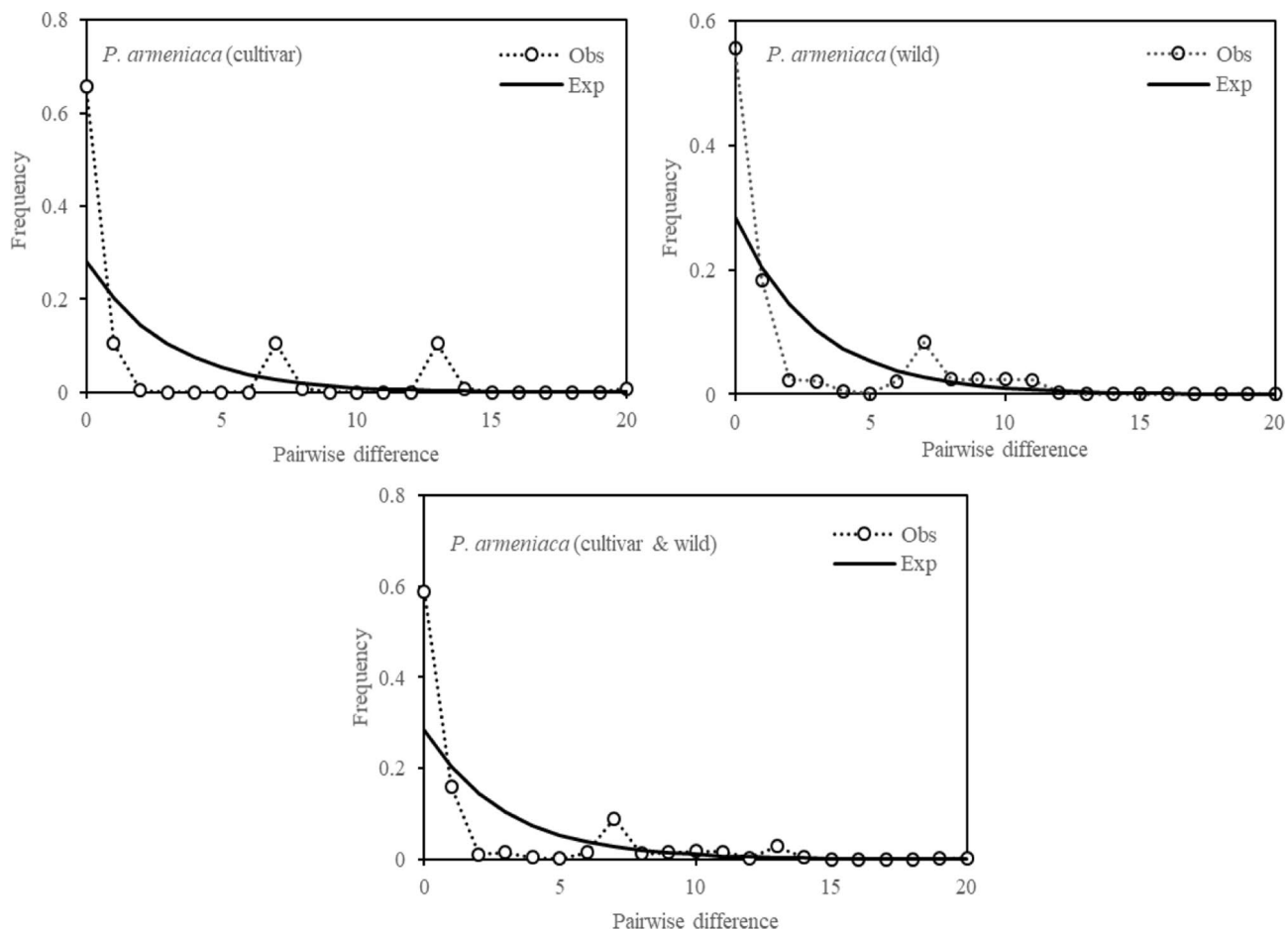
**Table 3.** Estimates of average gene diversity within populations ( $H_S$ ), total gene diversity ( $H_T$ ), interpopulation differentiation ( $G_{ST}$ ) and number of substitution types ( $N_{ST}$ ) for cpDNA haplotypes and ITS genotypes of *P. armeniaca*. Ns, not significant; \*,  $P < 0.05$ .

Source of variation	cpDNA					ITS				
	df	SS	VC	PV (%)	$F_{ST}$	df	SS	VC	PV (%)	$F_{ST}$
Among populations	11	31.699	0.209	16.28		11	23.777	0.068	4.38	
Within populations	95	102.086	1.075	83.72		112	166.844	1.490	95.62	
Total	106	133.785	1.284		0.1628*	123	190.621	1.558		0.0297*

**Table 4.** Analyses of molecular variance (AMOVA) of cpDNA haplotypes and ITS genotypes for populations of *P. armeniaca*. df, degrees of freedom; SS, sum of squares; VC, variance components; PV, percentage of variation;  $F_{ST}$ , genetic differentiation; \* $P < 0.001$ .

16.28%; ITS: 2.97%), which was consistent with previous results based on simple sequence repeat markers<sup>56</sup>. The relatively high genetic diversity also confirmed the Tianshan Mountains as the origin center of cultivated apricot<sup>56</sup>. The limited informative mutation sites among the ITS genotypes led to very little resolution for the construction of genotype relationships (Figure S2), suggesting rapid intraspecific differentiation in the recently derived species *P. armeniaca*, similar to the results found in *S. sinomontana*<sup>32</sup>.

The genetic backgrounds of the related species had the same genotypes (T2, T3, T11, T25, T8, T9, and T12) as *P. armeniaca*, indicating that they are associated with *P. armeniaca* through continuous and extensive gene



**Figure 4.** Mismatch distribution analysis of *P. armeniaca* based on overall gene pool of cpDNA dataset.

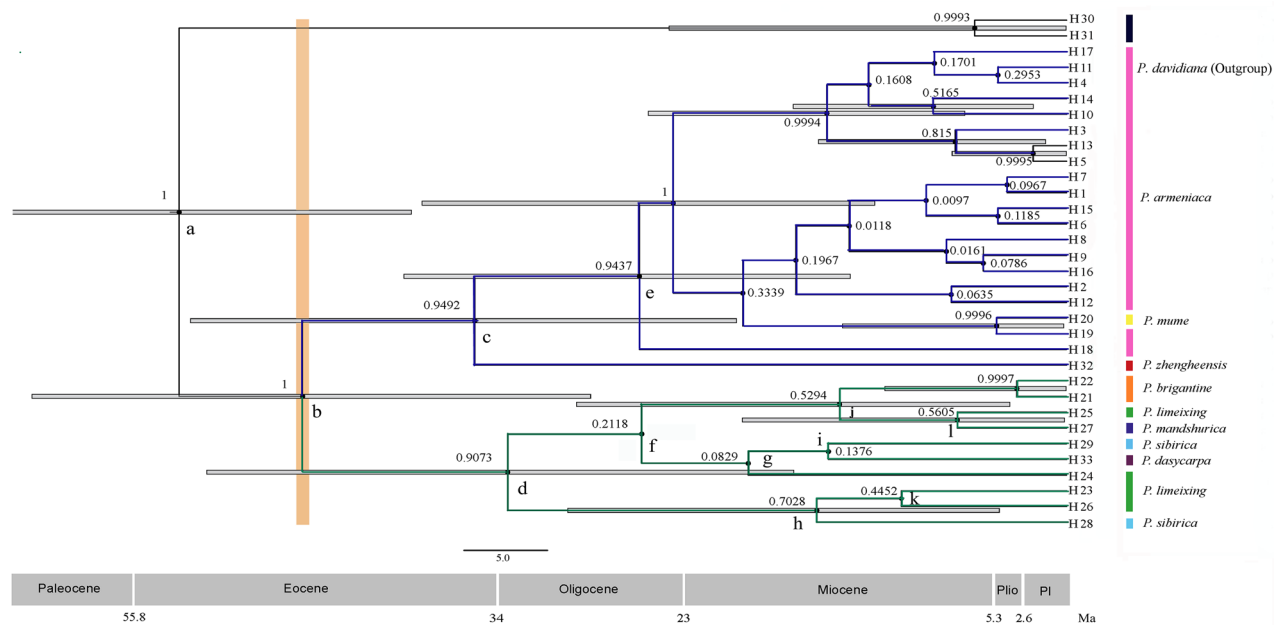
Populations		Tajima's $D$ -test		Fu's $F_S$ -test		Mismatch distribution			
		$D$	$P$	$F_S$	$P$	SSD	$P$	HRI	$P$
cpDNA	Overall	-2.417	0.000*	-6.350	0.025*	0.232	0.150	0.218	0.600
	<i>P. armeniaca</i> (cultivated)	-1.933	0.003*	1.289	0.772	0.038	0.250	0.356	0.500
	<i>P. armeniaca</i> (wild)	-2.272	0.001*	-5.825	0.021*	0.037	0.070	0.179	0.880
ITS	Overall	-1.193	0.106	-8.587	0.010*	0.002	0.660	0.023	0.700
	<i>P. armeniaca</i> (cultivated)	-1.420	0.155	-4.951	0.024*	0.015	0.020	0.063	0.310
	<i>P. armeniaca</i> (wild)	-0.966	0.041*	-12.223	0.000*	0.002	0.810	0.017	0.880

**Table 5.** Neutrality tests (Tajima's  $D$  and Fu's  $F_S$ ) and mismatch distribution analysis for the *P. armeniaca* based on the cpDNA and ITS dataset. \* $P < 0.05$ .

flow<sup>57</sup>. Liu et al.<sup>58</sup> concluded that *P. sibirica* was divided into two groups based on microsatellite markers, one of which may have undergone gene exchange with *P. armeniaca*, further verifying our results. In addition, the authors found an extensively mixed genetic background in the germplasm of cultivated apricots in China.

This study indicated that the cultivated and wild populations of *P. armeniaca* had the same ancestral haplotype, G1. The haplotypes of the CAG, EG and NCG populations were mixed with the haplotypes/genotypes of the large wild populations (*P. armeniaca*). According to coalescent theory<sup>59</sup>, chloroplast haplotype G1, which was widely distributed and located in the center of the chloroplast network (Fig. 2), should be considered the oldest haplotype. The Kashgar, Hotan and Aksu oasis areas around the Tarim Basin in the southern part of the Xinjiang Uygur Autonomous Region of China are the main apricot-producing areas and contain the greatest abundance of apricot cultivars. There is only one mountain between southern Xinjiang and the Ili Valley, and there are several corridors between the northern and southern Tianshan Mountains. Therefore, the apricots cultivated in Xinjiang, southern Tianshan Mountains (CAG), most likely evolved from the spread of wild apricots in the Ili River Valley. Liu et al.<sup>58</sup> argued that apricots have experienced at least three domestication events, giving rise to apricots in Europe (the United States and continental Europe), southern Central Asia (Turkmenistan,





**Figure 5.** Best-derived chronogram of *Prunus* spp. based on chloroplast DNA. Gray bars represent 95% highest posterior density intervals. The numbers on the branches are posterior probabilities (PP > 0.9). Plio, Pliocene; PI, Pleistocene; numbers above the branches indicate Bayesian posterior probabilities.

Afghanistan, and India) and China, with both ancient gene flow and recent gene mixing occurring. Central Asia harbors the highest diversity of wild apricots, with genetically differentiated populations that may have resulted from population isolation in glacial refugia<sup>58</sup>. In this study, the nucleotide diversity and haplotype/genotype diversity of wild apricots (DZG accessions) were higher than those of CAG and NCG accessions. These findings are reasonable from a historical perspective, as there was extensive cultural contact along the Silk Road from 207 BCE to 220 CE<sup>60</sup>. Therefore, historical and commercial influences may have contributed to the development of this unique species of cultivated apricot.

The theory and method of phylogeography can reveal the historical dynamics of species or populations, such as expansion, differentiation, isolation, migration and extinction<sup>29</sup>. It is of great significance for us to understand the origin of species and the evolution of geographical patterns, and to better protect existing biodiversity. Phylogeographic studies have shown that ancient haplotypes and high genetic diversity can be used to identify refuges<sup>1,61</sup>. Populations in refugia usually display more genetic diversity and exclusive haplotypes than migratory populations<sup>14</sup>. Many scholars<sup>17,19,24</sup> have suggested that the complex geographic history of Northwest China may have provided refuges for species during glacial periods. By combining two markers, we showed that all the wild populations of apricots distributed in the Ili River Valley contained ancestral haplotypes/genotypes and had high genetic diversity (Table 2). These populations were located in areas considered glacial refugia for *P. armeniaca*, which appears to be a relic of Quaternary glaciation. The region provides a suitable climate for the biological community and protects the genetic diversity of *P. armeniaca*. Climatic changes during Pleistocene glacial-interglacial cycles had a dramatic effect on species distribution ranges, causing migration and/or extinction of populations, followed by periods of isolation, divergence and subsequent expansion<sup>14</sup>. Both neutrality tests and mismatch distribution analysis based on the cpDNA and ITS datasets suggested recent range or demographic expansion of wild populations of *P. armeniaca*. We estimated that the recent demographic expansion of the wild populations of *P. armeniaca* occurred 16.53 kyr ago, that is, at the end of the LGM<sup>49</sup>.

The selected taxon-sampling and fossil calibration strategies will influence the age estimation<sup>62–63</sup>. Due to the lack of fossil evidence for *P. armeniaca*, we used the peach-apricot divergence time as the calibration point. In this study, we tried to use a cpDNA dataset to estimate divergence time, and the effect was acceptable. However, compared with the median ages estimated by Chin et al.<sup>50</sup> (mean age of 31.1 Ma), the divergence time estimates in this study should be interpreted with caution because the limited coverage and low number of calibration points may lead to an overly high divergence time estimate for *P. sibirica* (mean age of 33.43 Ma).

## Conclusion

Based on cpDNA and ITS data, the haplotype/genotype diversity of wild apricot populations distributed in the Ili River Valley was relatively high, and the haplotype/genotype diversity of DZGcmd and DZGyn populations was greater than that of other populations. *P. armeniaca* exhibits genealogical structure. Affected by the Quaternary glaciation of the Pleistocene, the Ili River Valley in Northwest China served as a glacial refugium for *P. armeniaca*, providing the species with a suitable climate and preserving its genetic diversity. During the interglacial period, the species underwent a recent expansion in the face of favorable climatic and environmental

conditions. Apricots originated during the middle Eocene, and the cultivated apricot in Xinjiang originated from apricots in the Ili River Valley in Northwest China.

**Research involving plants.** The experimental research and field studies on plants in this work comply with the IUCN Policy Statement on Research Involving Species at Risk of Extinction and the Convention on the Trade in Endangered Species of Wild Fauna and Flora.

Received: 6 February 2021; Accepted: 15 June 2021

Published online: 01 July 2021

## References

- Meng, H. H. & Zhang, M. L. Diversification of plant species in arid Northwest China: species-level phylogeographical history of *Lagochilus Bunge ex Benth* (Lamiaceae). *Mol. Phylogenet. Evol.* **68**, 398–409. <https://doi.org/10.1111/jse.12088> (2015).
- Pennington, R. T. *et al.* Contrasting plant diversification histories within the Andean biodiversity hotspot. *Proc. Natl. Acad. Sci. USA* **107**, 13783–13787. <https://doi.org/10.1073/pnas.1001317107> (2010).
- Hughes, C. & Eastwood, R. Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. *Proc. Natl. Acad. Sci. USA* **103**, 10334–10339. <https://doi.org/10.1073/pnas.0601928103> (2006).
- Johansson, U. S. *et al.* Build-up of the Himalayan avifauna through immigration: a biogeographical analysis of the *Phylloscopus* and *Seicercus* warblers. *Evolution* **61**, 324–333. <https://doi.org/10.1111/j.1558-5646.2007.00024.x> (2007).
- Hughes, C. E. & Atchison, G. W. The ubiquity of alpine plant radiations: from the Andes to the Hengduan Mountains. *New Phytol.* **207**, 275–282. <https://doi.org/10.1111/nph.13230> (2015).
- Lagomarsino, L. P., Condamine, F. L., Antonelli, A., Mulch, A. & Davis, C. C. The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytol.* **210**, 1430–1442. <https://doi.org/10.1111/nph.13920> (2016).
- Ebersbach, J. *et al.* In and out of the Qinghai-Tibet Plateau: divergence time estimation and historical biogeography of the large arctic-alpine genus *Saxifraga* L. *J. Biogeogr.* **44**, 900–910. <https://doi.org/10.1111/jbi.12899> (2017).
- Zhang, J. Y. & Zhang, Z. In *Flora of Chinese Fruit Trees* 61–62 (China Forestry Press, 2003).
- Su, Z., Zhang, M. & Sanderson, S. C. Chloroplast phylogeography of *Helianthemum songaricum* (Cistaceae) from northwestern China: implications for preservation of genetic diversity. *Conserv. Genet.* **12**, 1525–1537. <https://doi.org/10.1007/s10592-011-0250-9> (2011).
- Xie, K. Q. & Zhang, M. L. The effect of Quaternary climatic oscillations on *Ribes meyeri* (Saxifragaceae) in northwestern China. *Biochem. Syst. Ecol.* **50**, 39–47. <https://doi.org/10.1016/j.bse.2013.03.031> (2013).
- Salvi, D., Schembri, P., Sciberras, A. & Harris, D. Evolutionary history of the maltese wall lizard *Podarcis filfolensis*: insights on the ‘Expansion–Contraction’ model of Pleistocene biogeography. *Mol. Ecol.* **23**, 1167–1187. <https://doi.org/10.1111/mec.12668> (2014).
- Liu, J. Q., Sun, Y. S., Ge, X. J., Gao, L. M. & Qiu, Y. X. Phylogeographic studies of plants in China: advances in the past and directions in the future. *J. Syst. Evol.* **50**, 267–275. <https://doi.org/10.1111/j.1759-6831.2012.00214.x> (2012).
- Hewitt, G. The genetic legacy of the quaternary ice ages. *Nature* **405**, 907–913. <https://doi.org/10.1038/35016000> (2000).
- Hewitt, G. M. The structure of biodiversity—insights from molecular phylogeography. *Front. Zool.* **1**, 1–16. <https://doi.org/10.1186/1742-9994-1-4> (2004).
- Willis, K. J. & Niklas, K. J. The role of quaternary environmental change in plant macroevolution: the exception or the rule?. *Philos. Trans. R. Soc. Lond. B* **359**, 159–172. <https://doi.org/10.1098/rstb.2003.1387> (2004).
- Schmitt, T. Molecular biogeography of Europe: pleistocene cycles and postglacial trends. *Front. Zool.* **4**, 11. <https://doi.org/10.1186/1742-9994-4-11> (2007).
- Shen, L., Chen, X. Y. & Li, Y. Y. Glacial refugia and postglacial recolonization patterns of organisms. *Acta Ecol. Sin.* **22**, 1983–1990. <https://doi.org/10.1088/1009-1963/11/5/313> (2002).
- Schonswetter, P., Popp, M. & Brochmann, C. Rare arctic-alpine plants of the European Alps have different immigration histories: the snow bed species *Minuartia biflora* and *Ranunculus pygmaeus*. *Mol. Ecol.* **15**, 709–720. <https://doi.org/10.1111/j.1365-294X.2006.02821.x> (2006).
- Guo, Y. P., Zhang, R., Chen, C. Y., Zhou, D. W. & Liu, J. Q. Allopatric divergence and regional range expansion of *Juniperus sabina* in China. *J. Syst. Evol.* **48**, 153–160. <https://doi.org/10.1111/j.1759-6831.2010.00073.x> (2010).
- Jaramillo-Correa, J. P., Beaulieu, J. & Bousquet, J. Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental North American conifer. *Mol. Ecol.* **13**, 2735–2747. <https://doi.org/10.1111/j.1365-294X.2004.02258.x> (2004).
- Afzal-Rafii, Z. & Dodd, R. S. Chloroplast DNA supports a hypothesis of glacial refugia over postglacial recolonization in disjunct populations of black pine (*Pinus nigra*) in western Europe. *Mol. Ecol.* **16**, 723–736. <https://doi.org/10.1111/j.1365-294X.2006.03183.x> (2007).
- Anderson, L., Hu, F., Nelson, D., Petit, R. & Paige, K. Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *Proc. Natl. Acad. Sci. USA* **103**, 12447–12450. <https://doi.org/10.1073/pnas.0605310103> (2006).
- Volkova, P. A., Burlakov, Y. A. & Schanzer, I. A. Genetic variability of *Prunus padus* (Rosaceae) elaborates “a new Eurasian phylogeographical paradigm”. *Plant Syst. Evol.* **306**, 1–9. <https://doi.org/10.1007/s00606-020-01644-0> (2020).
- Xu, Z. & Zhang, M. L. Phylogeography of the arid shrub *Atraphaxis frutescens* (Polygonaceae) in northwestern China: evidence from cpDNA sequences. *J. Hered.* **106**, 184–195. <https://doi.org/10.1093/jhered/esu078> (2015).
- Rehder, A. *Manual of Cultivated Trees and Shrubs Hardy in North America, Exclusive of the Subtropical and Warmer Temperate Regions* 345–346 (Macmillan, 1927).
- Zhebentyayeva, T. N., Ledbetter, C., Burgos, L., & Llácer, G. *Fruit Breeding* 415–458 (Springer, 2012).
- Zhebentyayeva, T. N., Reighard, G. L., Gorina, V. M. & Abbott, A. G. Simple sequence repeat (SSR) analysis for assessment of genetic variability in apricot germplasm. *Theor. Appl. Genet.* **106**, 435–444. <https://doi.org/10.1007/s00122-002-1069-z> (2003).
- Schaal, B. A., Hayworth, D. A., Olsen, K. M., Rauscher, J. T. & Smith, W. A. Phylogeographic studies in plants: problems and prospects. *Mol. Ecol.* **7**, 465–474. <https://doi.org/10.1046/j.1365-294x.1998.00318.x> (1998).
- Avise, J. C. Phylogeography: retrospect and prospect. *J. Biogeogr.* **36**, 3–15. <https://doi.org/10.1111/j.1365-2699.2008.02032.x> (2009).
- Poudel, R. C., Möller, M., Li, D. Z., Shah, A. & Gao, L. M. Genetic diversity, demographical history and conservation aspects of the endangered yew tree *Taxus contorta* (syn. *Taxus fuana*) in Pakistan. *Tree Genet. Genom.* **10**, 653–665. <https://doi.org/10.1007/s11295-014-0711-7> (2014).
- Dutech, C., Maggia, L. & Joly, H. Chloroplast diversity in *Vouacapoua americana* (Caesalpinaceae), a neotropical forest tree. *Mol. Ecol.* **9**, 1427–1432. <https://doi.org/10.1046/j.1365-294x.2000.01027.x> (2000).
- Li, Y. *et al.* Rapid intraspecific diversification of the Alpine species *Saxifraga sinomontana* (Saxifragaceae) in the Qinghai-Tibetan Plateau and Himalayas. *Front. Genet.* **9**, 381. <https://doi.org/10.3389/fgene.2018.00381> (2018).

33. Zhang, Q. P. & Liu, W. S. Advances of the apricot resources collection, evaluation and germplasm enhancement. *Acta Hort. Sin.* **45**, 1642–1660. <https://doi.org/10.16420/j.issn.0513-353x.2017-0654> (2018).
34. Hu, Z. B. *et al.* Population genomics of pearl millet (*Pennisetum glaucum* (L.) R. Br.): comparative analysis of global accessions and Senegalese landraces. *BMC Genomics* **16**, 1048. <https://doi.org/10.1186/s12864-015-2255-0> (2015).
35. White, T. J., Bruns, T., Lee, S. & Taylor, J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protoc. Guide Methods Appl.* **18**, 315–322 (1990).
36. Dong, W. *et al.* ycf1, the most promising plastid DNA barcode of land plants. *Sci. Rep.* **5**, 8348. <https://doi.org/10.1038/srep08348> (2015).
37. Bortiri, E. *et al.* Phylogeny and systematics of *Prunus* (Rosaceae) as determined by sequence analysis of ITS and the chloroplast *trnL-trnF* spacer DNA. *Syst. Bot.* **26**, 797–807. <https://doi.org/10.1043/0363-6445-26.4.797> (2001).
38. Zhang, Q. Y. *et al.* Latitudinal adaptation and genetic insights into the origins of *Cannabis sativa* L. *Front Plant Sci.* **9**, 1876. <https://doi.org/10.3389/fpls.2018.01876> (2018).
39. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Sumo. Ser.* **41**, 95–98. <https://doi.org/10.1021/bk-1999-0734.ch008> (1999).
40. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680. <https://doi.org/10.1093/nar/22.22.4673> (1994).
41. Simmons, M. P. & Ochoterena, H. Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* **49**, 369–381. <https://doi.org/10.1080/10635159950173889> (2000).
42. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874. <https://doi.org/10.1093/molbev/msw054> (2016).
43. Clement, M., Posada, D. & Crandall, K. A. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* **9**, 1657–1659. <https://doi.org/10.1046/j.1365-294x.2000.01020.x> (2000).
44. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452. <https://doi.org/10.1093/bioinformatics/btp187> (2009).
45. Pons, O. & Petit, R. J. Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* **144**, 1237–1245. [https://doi.org/10.1016/S1050-3862\(96\)00162-3](https://doi.org/10.1016/S1050-3862(96)00162-3) (1996).
46. Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Ecol. Resour.* **10**, 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x> (2010).
47. Rogers, A. R. & Harpending, H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9**, 552–569 (1992).
48. Wolfe, K. H., Li, W. H. & Sharp, P. M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. USA* **84**, 9054–9058. <https://doi.org/10.1073/pnas.84.24.9054> (1987).
49. Wang, Z. *et al.* Phylogeography study of the Siberian apricot (*Prunus sibirica* L.) in Northern China assessed by chloroplast microsatellite and DNA makers. *Front. Plant Sci.* **8**, 1989. <https://doi.org/10.3389/fpls.2017.01989> (2017).
50. Chin, S. W., Shaw, J., Haberle, R., Wen, J. & Potter, D. Diversification of almonds, peaches, plums and cherries—Molecular systematics and biogeographic history of *Prunus* (Rosaceae). *Mol. Phylogenet. Evol.* **76**, 34–48. <https://doi.org/10.1016/j.ympev.2014.02.024> (2014).
51. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973. <https://doi.org/10.1093/molbev/mss075> (2012).
52. Yang, J., Vazquez, L., Feng, L., Liu, Z. & Zhao, G. Climatic and soil factors shape the demographical history and genetic diversity of a deciduous oak (*Quercus liaotungensis*) in Northern China. *Front. Plant Sci.* **9**, 1534. <https://doi.org/10.3389/fpls.2018.01534> (2018).
53. Zhang, X., Shen, S., Wu, F. & Wang, Y. Inferring genetic variation and demographic history of *Michelia yunnanensis* Franch (Magnoliaceae) from chloroplast DNA sequences and microsatellite markers. *Front. Plant Sci.* **8**, 583. <https://doi.org/10.3389/fpls.2017.00583> (2017).
54. Li, M., Zhao, Z. & Miao, X. J. Genetic variability of wild apricot (*Prunus armeniaca* L.) populations in the Ili Valley as revealed by ISSR markers. *Genet. Resour. Crop Evol.* **60**, 2293–2302. <https://doi.org/10.1007/s10722-013-9996-x> (2013).
55. Li, M., Hu, X., Miao, X. J., Xu, Z. & Zhao, Z. Genetic diversity analysis of wild apricot (*Prunus armeniaca*) populations in the Ili Valley as revealed by SRAP markers. *Acta Hort. Sin.* **43**, 1980–1988. <https://doi.org/10.16420/j.issn.0513-353x.2016-0156> (2016).
56. Hu, X., Zheng, P., Ni, B., Miao, X. & Li, M. Population genetic diversity and structure analysis of wild apricot (*Prunus armeniaca* L.) revealed by SSR markers in the Tien-Shan mountains of China. *Pak. J. Bot.* **50**, 609–615 (2018).
57. Decroocq, S. *et al.* New insights into the history of domesticated and wild apricots and its contribution to Plum pox virus resistance. *Mol. Ecol.* **25**, 4712–4729. <https://doi.org/10.1111/mec.13772> (2016).
58. Liu, S. *et al.* The complex evolutionary history of apricots: species divergence, gene flow and multiple domestication events. *Mol. Ecol. Notes* **28**, 5299–5314. <https://doi.org/10.1111/mec.15296> (2019).
59. Posada, D. & Crandall, K. A. Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol. Evol.* **16**, 37–45. [https://doi.org/10.1016/S0169-5347\(00\)02026-7](https://doi.org/10.1016/S0169-5347(00)02026-7) (2001).
60. Boulnois, L. *Silk Road: Monks, Warriors & Merchants on the Silk Road* 115–165 (WW Norton & Co Inc, 2004).
61. Zhao, C., Wang, C. B., Ma, X. G., Liang, Q. L. & He, X. J. Phylogeographic analysis of a temperate-deciduous forest restricted plant (*Bupleurum longiradiatum* Turcz.) reveals two refuge areas in China with subsequent refugial isolation promoting speciation. *Mol. Phylogenet. Evol.* **68**, 628–643. <https://doi.org/10.1016/j.ympev.2013.04.007> (2013).
62. Ebersbach, J., Schnitzler, J., Favre, A. & Muellner-Riehl, A. N. Evolutionary radiations in the species-rich mountain genus *Saxifraga* L. *BMC Evol. Biol.* <https://doi.org/10.1186/s12862-017-0967-2> (2017).
63. Favre, A. *et al.* The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biol. Rev.* **90**, 236–253. <https://doi.org/10.1111/brv.12107> (2015).

## Acknowledgements

We are grateful to K.L. for the useful discussions and insightful comments and to L.Q.L., Q.P.Z., W.Q.Z. and G.Q.F. for the assistance in collecting samples.

## Author contributions

W.W.L., and K.L. conceived the study. L.Q.L., Q.P.Z., W.Q.Z. and G.Q.F. contributed to the sampling. W.W.L. collected and analyzed the data. W.W.L. and K.L. wrote the manuscript. All authors read and approved the final version of the manuscript.

## Funding

This work was supported by the National Key Research and Development Program (grant number 2016YFC0501504), the Xinjiang Uygur Autonomous Region Horticulture Key Discipline Fund (grant number 2016-10758-3), the Project funded by China Postdoctoral Science Foundation (grant number 2021M693898) and the Xinjiang Agricultural University Crop science postdoctoral research station. The authors would like to thank the different research institutions, scientists, and breeders involved in this work and the company American Journal Experts (AJE) for providing English editorial assistance.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-93050-w>.

**Correspondence** and requests for materials should be addressed to K.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021