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

Wen-Wen Li¹, Li-Qiang Liu¹, Qiu-Ping Zhang², Wei-Quan Zhou¹, Guo-Quan Fan³ & Kang Liao¹*

To clarify the phytogeography 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 conservation¹. 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 strategies²,³. The origins of mountain biodiversity are complex and may include the immigration of preadapted lineages⁴–⁵, in situ diversification⁶, or the continuation of ancestral lineages⁷. 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, China⁸. However, there have been few studies on the phylogeography of plant species in the arid region of Northwest China⁹–¹⁰.

In recent decades, the method of combining molecular data with paleoclimatic and geographical evidence has been effective in the study of systematic geography¹¹–¹². 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 Hemisphere¹³–¹⁴. Especially during the Quaternary glacial period, species in the ice-free zone were mainly affected by the cold and dry climate¹⁵. Climatic fluctuations cause the distribution of species to shrink and expand¹⁶, and cold and dry climates prompt plant and animal species to retreat to refugia.

¹College of Horticulture and Forestry, Xinjiang Agricultural University, Urumqi, Xinjiang, China. ²Xiongyue National Germplasm Resources Garden of the Liaoning Institute of Pomology, Xiongyue, Shenyang, China. ³Luntai National Fruit Germplasm Resources Garden of Xinjiang Academy of Agricultural Sciences, Luntai, Xinjiang, China. ⁴email: liaokang01@163.com
which provided shelter. As temperatures rose after the glacial period, species underwent population expansion, changes that left a genetic signature in their current populations. Although no unified glacial sheet was present in Asia, palaeodata suggest that the distribution ranges of woody species in this region were similarly changed by Quaternary climatic oscillations. Based on the geographical distribution of genetic variation, the glacial refuges and postglacial re-vegetation routes of most woody plants are roughly consistent with fossil evidence. 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. Volkova 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. The complex geographic history in the area may have provided a refuge for species in the last glacial period. Xu et al. proposed two possible independent glacial refugia in Northwest China: the Ili River Valley and the Northern Junggar Basin. Su et al. 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). Almost all cultivated apricots originated from *Prunus armeniaca*. 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. 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. Organelle markers could provide powerful tools for studying the phylogeography and migratory footprints of species. Parental genetic markers are often combined with single-parent organelle markers for population genetics studies. cpDNA and ITS sequence variations have been very effective in revealing the glacial refuges of plants. cpDNA lineages usually show the unique geographical distribution and evolutionary history of natural populations and are therefore widely used in systematic geography studies. Li et al. 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 syriaca, NAG; Prunus mandshurica, LX; Prunus sasanqua, ZN; Prunus nema, ECG; Prunus zhengheensis, ZHX*; *Prunus limeixing, LMX* 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 (DYGp population), Gongliu County (DZGgb and DZGgld populations) and Xinyuan County (DZGytx, DZGyxa and DZGyxz 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, Shaxanxi, 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.

#### DNA sequencing.

Total genomic DNA was extracted from the silica gel-dried leaf materials using a Plant Genomic DNA Kit (Tiangen Biotech, Beijing, China). 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. PCR was performed in a total volume of 25 μL that contained 1 μL DNA, 5.5 μL PCR mix, 16.5 μL double-distilled water and 1 μL 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 °C, 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) was used for
purification. The purified PCR products were sequenced on an ABI PRISM 3730XL DNA Analyzer (Applied Biosystems, Foster City, CA, USA).38.

**Data analysis.** We used BioEdit39 to view and manually correct the sequencing results. We first used CLUSTAL W40 to align the sequences and coded indels following the method of Simmons and Ochoterena41. Then, manual adjustments were made in MEGA ver. 7.0.2642 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. 1.2.143. ArcGIS ver. 10.2 (http://desktop.arcgis.com) software was used to create a haplotype geographical distribution map.

Haplotype/genotype diversity (Hd) and nucleotide diversity (π) were calculated using DnaSP ver. 5.10 software.44. The within-population gene diversity (Hs), gene diversity in all populations (HT), interpopulation differentiation (GST) and number of substitution types (NST) were calculated using PERMUT45 ver. 2.0. The last two indexes (GST and NST) were analyzed via permutation tests with 1000 permutations. When NST is greater than GST, it indicates the existence of genealogical geographic structure45. Analysis of molecular variance (AMOVA) was performed using Arlequin ver. 3.5.2.2 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 (HR1)46 and corresponding P values were calculated using Arlequin ver. 3.5.2.2.46. Neutrality tests based on Tajima’s D and Fu’s Fw were conducted to detect departures from the population equilibrium by Arlequin ver. 3.5.2.246. According to the formula of Rogers and Harpending47, T = r/2u, where “r” is the parameter value from the mismatch distribution model. In the formula u = μkg, “μ” is the base substitution rate (chloroplast angiosperms48: 1.1 × 10−9), “k” is the fragment length (cpDNA length after combination: 2062 bp), and “g” is the generation time (20 years)49. The outgroup was *P. davidiana*, and the time point of peach-apricot differentiation was used as the calibration point50. The BEAUti interface was used to create an input file for BEAST51, 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

| Population | Origin/location | Type | Latitude (°N) | Longitude (°E) | Altitude (m) |
|------------|-----------------|------|--------------|----------------|-------------|
| DZGhkemd  | Huocheng, Xinjiang | W   | 44.43        | 80.79          | 1187.6      |
| DZGkchy   | Huocheng, Xinjiang | W   | 44.44        | 80.79          | 1245.7      |
| DZGkchm   | Huocheng, Xinjiang | W   | 44.40        | 80.71          | 1244.1      |
| DZGyn     | Yining, Xinjiang  | W   | 44.12        | 81.62          | 1983.6      |
| DZGqgbl   | Gongliao, Xinjiang | W   | 43.25        | 82.86          | 1371.5      |
| DZGqgdl   | Gongliao, Xinjiang | W   | 43.23        | 82.75          | 1269.4      |
| DZGxyl    | Xinyuan, Xinjiang | W   | 43.54        | 83.44          | 1138.8      |
| DZGxyl    | Xinyuan, Xinjiang | W   | 43.50        | 83.70          | 1275.3      |
| DZGxyl    | 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        |

Table 1. Population code, sampling location, coordinates, altitude and types of *P. armeniaca* and relative species. W, wild; C, cultivars; O, out group.
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. mumme* 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. mumme* 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: 0.0013, 0.444, respectively) was higher than that of the CAG (0.0006, 0.239) and NCG (0.0002, 0.400) 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 structure and genealogical geography. In *P. armeniaca*, the gene diversity among populations (cpDNA: *Hs* = 0.499; ITS: *Hs* = 0.876) was higher than that within populations (cpDNA: *Hs* = 0.490; ITS: *Hs* = 0.794) (Table 3). A permutation test showed that *Nst* was significantly higher than *Gst* (cpDNA: *Nst* = 0.227 > *Gst* = 0.020; ITS: *Nst* = 0.126 > *Gst* = 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: *Fst* = 0.1628, *P* < 0.001; ITS: *Fst* = 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 and relatively little occurring among them (Table 4).
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. used cpDNA *trnL-trnF, rp16* and *nrDNA ITS* sequences to infer the evolutionary history of *S. simulans*. Yang et al. used cpDNA *psbA-trnH, trnL-trnF, ycf1*, and *matK* sequences to access the demographical history and genetic diversity of a Deciduous Oak (*Quercus liaotungensis*) in Northern China. Zhang et al. used cpDNA markers to successfully determine the genetic diversity, genetic structure, and demographic history of 7 *Michelia yunnanensis* populations. Many scholars 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. 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. 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: 16.28%; ITS: 2.97%).
16.28%; ITS: 2.97%), which was consistent with previous results based on simple sequence repeat markers. The relatively high genetic diversity also confirmed the Tianshan Mountains as the origin center of cultivated apricot. 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.

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.

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.

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   | 23.78 | 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.

**Species**

- **P. davidiana**
- **P. dasyacarpa**
- **P. mume**
- **P. armeniaca**
- **P. sibirica**
- **P. limeixing**
- **P. mandshurica**
- **P. zhengheensis**
flow. Liu et al. 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, 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. 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 4.** Mismatch distribution analysis of *P. armeniaca* based on overall gene pool of cpDNA dataset.

**Table 5.** Neutrality tests (Tajima’s *D* and Fu’s *F*_I*) and mismatch distribution analysis for the *P. armeniaca* based on the cpDNA and ITS dataset. *P* < 0.05.
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. 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. 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. 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 refugia. Populations in refugia usually display more genetic diversity and exclusive haplotypes than migratory populations. Many scholars 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. 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.

The selected taxon-sampling and fossil calibration strategies will influence the age estimation. 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. (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 DZGhcmd 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.
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
1. Meng, H. H. & Zhang, M. L. Diversification of plant species in arid Northwest China: species-level phylogeographical history of Lagochilus Bunge ex Bentham (Lamiaceae). Mol. Phylogenet. Evol. 68, 398–409. https://doi.org/10.1016/j.ympev.2015.05.002 (2015).
2. 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).
3. 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).
4. 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/10.1111/1558-5646.12002 (2007).
5. 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).
6. Lagomarsino, L. P., Condamine, F. L., Antonelli, A., Mulch, A. & Davis, C. C. The abiotic and biotic drivers of rapid diversification in Andean belflower (Campanulaceae). New Phytol. 210, 1430–1442. https://doi.org/10.1111/nph.13920 (2016).
7. 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, 906–910. https://doi.org/10.1111/jbi.12899 (2017).
8. Zhang, J. Y. & Zhang, Z. In Flora of Chinese Fruit Trees 61–62 (China Forestry Press, 2003).
9. Su, Z., Zhang, M. & Sanderson, S. C. Chloroplast phylogeography of apricots in the Ili River Valley in Northwest China. Tree Genet. Genomes 5, 184–195. https://doi.org/10.1007/s12267-009-0077-9 (2009).
10. Xie, K. Q. & Zhang, M. L. The effect of Quaternary climatic oscillations on Middle Eocene, and the cultivated apricot in Xinjiang originated from apricots in the Ili River Valley in Northwest China.
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.

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.

References

33. Zhang, Q. P. & Liu, W. S. Advances of the apricot resources collection, evaluation and germplasm enhancement. Acta Hortic. Sin. 45, 1642–1660. https://doi.org/10.14202/j.issn.0531-353x.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/10.1609/363-6445.26.4.797 (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/10.1016/bk-1999-0734.98008 (1999).
40. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of 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/10.1080/1063515990173889 (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, 1415–1424. 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/10.1093/genetics/144.4.1237 (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.84.21.9054 (1987).
49. Wang, Z. et al. Phylogeographic 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. Phylogen. 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. Trends Ecol. Evol. 29, 1969–1973. https://doi.org/10.1016/j.tree.2013.10.005 (2014).
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 SSR 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 L.) populations in the Ili Valley as revealed by SSRAP markers. Acta Hortic. 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. Biol. Evol. 28, 5299–5314. https://doi.org/10.1093/molbev/msr159 (2011).
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 (2001).
60. Bouchon, 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. Phylogen. Evol. 68, 628–643. https://doi.org/10.1016/j.ympev.2013.04.007 (2013).
62. Ebersbach, J., Schnitzler, J., Favre, A. & Mueller-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).
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.

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