CSIRO PUBLISHING

Australian Journal of Botany, 2012, **60**, 20–31
http://dx.doi.org/10.1071/BT11055

Phylogeography of the rare *Gymnocarpos przewalskii* (Caryophyllaceae): indications of multiple glacial refugia in north-western China

S. M. Ma[A,B], *M. L. Zhang*[A,C,E] *and S. C. Sanderson*[D]

[A]Key Laboratory of Biogeography and Bioresource in Arid Land, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China.
[B]Graduate University of the Chinese Academy of Sciences, Beijing 100049, China.
[C]Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China.
[D]Shrub Sciences Laboratory, Intermountain Research Station, Forest Service, U.S. Department of Agriculture, Utah 84606, USA.
[E]Corresponding author. Email: zhangml@ibcas.ac.cn

Abstract. We investigated the phylogeography of *Gymnocarpos przewalskii* Maxim. (Caryophyllaceae), a rare relictual shrub restricted to north-western China, in the context of Quaternary climate oscillations. Three cpDNA regions (*psb*A–*trn*H, *ycf*6–*psb*M and *rpl*32–*trn*L (UAG)) were sequenced for 160 individuals from 16 populations. High genetic diversity ($h_T = 0.930$, $h_S = 0.425$) and a significant phylogeographic structure ($N_{ST} > G_{ST}$, $P < 0.01$) were identified; 31 different cpDNA haplotypes were detected. Phylogenetic analyses showed that the haplotypes were clustered into five clades, consistent with their distributions in the following four geographic regions: the Tarim Basin, Hami Basin, the western Yumen of Gansu Province and an easternmost region, consisting of populations in the Wulate Rear Banner region in Inner Mongolia, the Jinta and Jingyuan regions in Gansu Province and the Zhongwei region in Ningxia. The existence of regional divergence was supported by AMOVA, which revealed that ~63% of variation was due to differences among the four geographic regions. Four independent glacial refugia were inferred, in the western Tarim Basin, Hami Basin, the Liuyuan region in western Gansu and the easternmost region mentioned. Population bottlenecks and postglacial recolonisation were identified in the northern Tarim Basin, western Yumen and the Jinta region in Gansu Province.

Received 16 February 2011, accepted 12 December 2011, published online 3 February 2012

Introduction

Climatic oscillations during the Quaternary, beginning ~2 million years ago, led to several glacial and interglacial cycles (Shackleton and Opdyke 1973). These paleoclimatic cycles resulted in repeated contractions and expansions of areas occupied by many organisms, and contributed to the restriction of species distributions to climate refugia during periods of maximum glaciation, and rapid range expansions when these periods ended (Hewitt 2000). Many studies of phylogeography have indicated that Pleistocene glaciations had a profound influence on the genetic structure of plant species throughout the northern hemisphere (Hewitt 2000). Pollen records have shown no evidence for glaciation in the desert regions of north-western China (Li 1998). However, in these areas, climate changes caused by the Tibetan Plateau uplift, particularly during Quaternary glaciations, together with geographical barriers (e.g. development of mountains and large deserts), have affected the distribution and evolution of many plant species (Ge *et al*. 2005, 2011).

Nevertheless, few phylogeographic studies have directly examined the evolutionary history of plant species restricted to the desert areas of north-western China. One case was a phylogeographic study in *Tetraena mongolica*, which indicated restricted gene flow and occasional long-distance dispersal for the species (Ge *et al*. 2011). According to Shen *et al*. (2002), the complex geographical history of north-western China probably provided refugia for species during the period of the last glacial maximum (LGM). For instance, absence of permafrost records suggests that large refugia existed in the Tarim Basin (Fig. 1) during the Middle and Late Pleistocene (Yang and Liu 2002). Studies of phylogeography of endemic plants from north-western China can therefore be of interest for the examination of genealogical lineages in the context of Quaternary climate oscillations, including range contractions, fragmentations, expansions and postglacial migrations from refugia.

Gymnocarpos przewalskii is mainly restricted to the deserts of north-western China. It has a geographically disjunct population distribution and differs morphologically by two unique features from the other nine species of *Gymnocarpos*, which are mainly distributed in western arid regions of the Cape Verde and Canary Islands (Oxelman *et al*. 2002). The

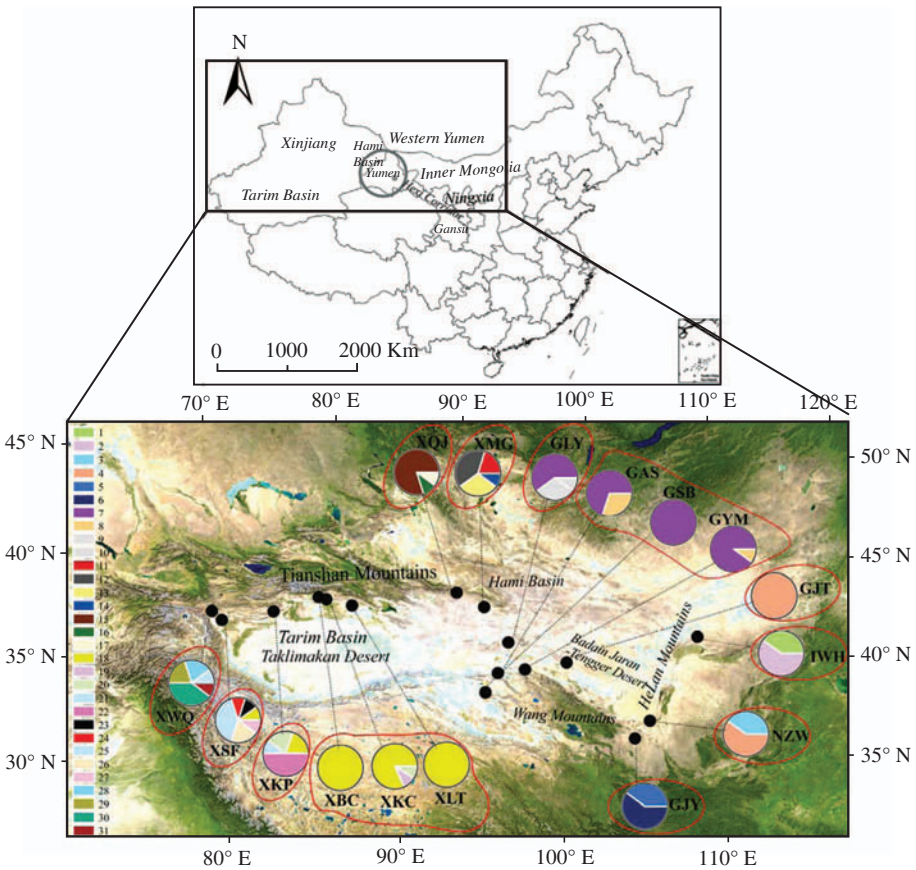

Fig. 1. Sampling localities and distribution of 31 chloroplast DNA haplotypes (labelled as 1–31), identified from 16 populations of *Gymnocarpos przewalskii* in desert areas of north-western China. Pie graphs reflect the frequency of each haplotype at these locations. Location abbreviations correspond to those listed in Table 1. The 12 groups of populations identified by SAMOVA are circled by the red dashed line, and correspond to those listed in Table 4. Geographic details of the area mentioned in the text are labelled.

presence of an inner pair of narrow lateral bracts below distal flowers, as well as a long style, more or less equal to the length of the sepals and much longer than the stamens, are the two unique features of *G. przewalskii*. The species mainly grows in semi-desert steppe, at stony or gravelly sites at up to ~2600-m altitude. Total vegetational coverage in *G. przewalskii* communities is often less than 10% (Wang 2005). It was defined as a rare species in the second rank of conservation priority by the China Plant Red Data Book (Fu 1992). In the past few decades, increasing human activities in its natural habitats have resulted in a serious loss of individuals, and caused habitat fragmentation and population isolation. Furthermore, the rate of seed set in *G. przewalskii* is low, less than 1%, especially under conditions of continuous aridity (e.g. monthly mean precipitation below 45 mm) during the flowering season (Chai *et al*. 2010). This frequently limits sexual propagation, leading to an increased proportion of clonal reproduction occurring by root formation on sand-covered branches. As a result, *G. przewalskii* may now be more vulnerable to decline and loss of genetic diversity. The species may also be an ideal candidate for molecular phylogeographic studies examining the evolutionary history of plants in the deserts of north-western China.

As a consequence, on the basis of 160 individuals belonging to 16 populations of *G. przewalskii*, collected across most of the species range, we implemented a phylogeographic study using three cpDNA spacers (*psb*A–*trn*H, *ycf*6–*psb*M and *rpl*32–*trn*L (UAG)). In plants, chloroplast DNA (cpDNA) is thought to evolve slowly, with low mutation (Li and Fu 1997) and recombination rates (Comes and Kadereit 1998). The maternal cpDNA lineages in natural populations often display distinct geographic distributions (Avise 2000), which can be informative about the evolutionary history of the species, and non-coding regions of cpDNA have been used successfully in phylogeographic studies (Dutech *et al*. 2000; Raspé *et al*. 2000). Available evidence also suggests that cpDNA sequence variation can be highly effective in revealing glacial refugia and postglacial expansion patterns in plants (Guo *et al*. 2010; Vidal-Russell *et al*. 2011). Here, our specific goals were to address the following questions: (1) how is the cpDNA variation geographically structured and is there phylogeographic structure in *G. przewalskii*; are there any historically isolated and genetically distinct lineages resulting from climate changes in north-western China; (2) are there independent glacial refugia consistant with the distributions of haplotype lineages; (3) is there evidence of recent expansion from these refugial sites

during more favourable climatic conditions; and (4) what strategies should be proposed to conserve the identified ancient haplotype lineages of *G. przewalskii*?

Materials and methods

Plant material

For this research, we conducted extensive field investigation from June 2008 to July 2010. Some populations of *G. przewalskii* reported in herbaria, such as those in Alashan left Banner of Inner Mongolia (Voucher specimen WUK 0346479), and Wushi (PE8420) and Aksu (PE8420) of the north-western Tarim Basin, have become extinct, probably because of frequent human activities in its natural habitats. In north-western Inner Mongolia, only one extant northernmost population, IWH, could be collected. In all, leaf samples from 160 individuals of *G. przewalskii* were collected from 16 populations covering the majority of the species range. These included a population, GJY, in the easternmost Hexi corridor of Gansu Province, where no specimen had been reported previously (Table 1, Fig. 1).

The majority of the investigated populations of *G. przewalskii* are reduced in size, especially GJT, XKC and XKP, where totals of only 10, 12 and 13 individuals, respectively, were found. For each population, with a few exceptions, 10 individuals spaced at least 5 m apart were sampled, and the fresh leaves were dried immediately in silica gel.

DNA extraction, PCR amplifications and sequencing

Total genomic DNA was extracted following the modified CTAB protocol (Doyle and Doyle 1987). Seven cpDNA regions (Table 2) were tested to detect possible intraspecific variation in individuals from different populations. Three spacers: *psb*A–*trn*H, *ycf*6–*psb*M and *rpl*32–*trn*L (UAG), containing the most polymorphic sites, were chosen for the full study. The DNA amplification profile was 94°C for 2 min, followed by 30 cycles of 94°C for 30 s, annealing at 53°C (*psb*A–*trn*H), 54°C (*ycf*6–*psb*M) or 53°C (*rpl*32–*trn*L (UAG)) for 30 s, 72°C for 90 s, and an additional extension in 72°C for 10 min.

PCR products were purified on 1.5% low-melting agarose gels. The desired DNA band was recovered with a UNIQ-10 kit (Shanghai Bioengineering, Shanghai, China), and sequenced in both directions in an ABI PRISM 3730 xl automated sequencer, following the manufacturer's protocol.

Data analysis

Sequences were edited and assembled with SeqMan v. 3.0 (Doyle and Doyle 1987), aligned using ClustalX v.1.83

Table 1. Localities sampled, and the number of total haplotypes, and estimated genetic diversity for private haplotypes in each population

Percentages in parentheses indicate the number of private haplotypes divided by the total number of haplotypes in each population. IM, Inner Mongolia; NX, Ningxia; GS, Gansu; XJ, Xinjiang; H_d, the haplotype diversity; Π, mean number of pairwise differences; and Π_n, nucleotide diversity

Population and location	Latitude (N)	Longitude (E)	Altitude (m)	No. of haplotypes	No. of private haplotypes	H_d	Π	Π_n
IM, Wulate Rear Banner, IWH	41°40.995′	108°30.968′	1025	2	2 (100%)	0.5333	24.0000	0.0132
NX, Zhongwei, NZW	37°29.032′	105°12.003′	1128	2	1 (50%)	0.5333	5.3333	0.0030
GS, Jingyuan, GJY	36°34.041′	104°12.990′	904	2	2 (100%)	0.5333	24.0000	0.0131
GS, Jinta, GJT	40°18.013′	99°30.004′	1180	1	0 (0%)	–	–	–
GS, Yumen, GYM	39°47.996′	96°42.123′	2312	2	0 (0%)	0.2000	1.20000	0.0007
GS, Subei, GSB	39°29.460′	94°53.008′	2213	1	0 (0%)	–	–	–
GS, Akesai, GAS	36°30.101′	104°11.892′	2736	2	0 (0%)	0.4667	2.8000	0.0016
GS, Liuyuan, GLY	39°5.966′	95°24.321′	1789	3	2 (67%)	0.6000	9.6667	0.0054
XJ, Miaoergou, XMG	43°5.005′	93°34.902′	1173	4	4 (100%)	0.7778	13.2000	0.0073
XJ, Qijiaojing, XQJ	43°20.512′	91°23.791′	1273	3	3 (100%)	0.3778	2.8000	0.0016
XJ, Luntai, XLT	41°46.023′	84°14.990′	979	1	0 (100%)	–	–	–
XJ, Kuche, XKC	41°48.224′	82°24.008′	1179	2	1 (50%)	0.2000	5.0000	0.0027
XJ, Baicheng, XBC	41°49.402′	81°50.381′	1260	1	0 (0%)	–	–	–
XJ, Keping, XKP	40°33.003′	79°4.601′	1175	4	3 (75%)	0.7333	11.2222	0.0061
XJ, Shufu, XSF	39°25.410′	75°50.131′	1359	6	5 (84%)	0.8889	17.1333	0.0094
XJ, Wuqia, XWQ	39°41.990′	75°06.003′	2198	5	4 (80%)	0.8222	15.6000	0.0086

Table 2. Chloroplast DNA regions surveyed before the study for population-level variation within *Gymnocarpos przewalskii*, showing primer sequences for PCR amplification and sequencing, and the source

F, forward; R, reverse

Region	Sequence 5′–3′	Source
*trn*S–*trn*G	F: GCCGCTTTAGTCCACTCAGC; R: GAACGAATCACACTTTTACCAC	Hamilton (1999)
*rpl*32–*trn*L (UAG)	F: CAGTTCCAAAAAAACGTACTTC; R: CTGCTTCCTAAGAGCAGCGT	Shaw *et al*. (2007)
*ndh*F–*rpl*32	F: GAAAGGTATKATCCAYGMATATT; R: CCAATATCCCTTYYTTTTCCAA	Shaw *et al*. (2007)
*psb*A–*trn*H	F: GTTATGCATGAACGTAATGCTC; R: CGCGCATGGTGGATTCACAATCC	Sang *et al*. (1997)
*ycf*6–*psb*M	F: TGGATATAGTAAGTCTYGCTTGGGC; R: AGTGCATGGAGTCTYGCTAGG	Shaw *et al*. (2007)
*trn*Q–*rps*16	F: GCGTGGCCAAGYGGTAAGGC; R: GTTGCTTTYTACCACATCGTTT	Sang *et al*. (1997)
*rpo*B–*trn*C (GCA)	F: CKACAAAAYCCYTCRAATTG; R: CACCCRGATTYGAACTGGGG	Shaw and Small (2005)

(Thompson *et al*. 1997), and refined by visual inspection. Molecular-diversity indices, including haplotype diversity (H_d), mean number of pairwise differences (Π; Tajima 1983) and nucleotide diversity (Π_n; mean number of pairwise differences per site; Nei 1987), were estimated using ARLEQUIN v.3.0 (Excoffier *et al*. 2005), for each population. DnaSP v.5.0 (Rozas *et al*. 2003) was used to identify the unique cpDNA haplotypes from all individuals. The geographic distribution of detected haplotypes was mapped using ArcMap 9.2 (ESRI, Redlands, CA, USA).

The distribution range of *G. przewalskii* was divided into four regions, based on the physico-geographical characteristics of north-western China, and also in accordance with phylogenetic analyses. The Tarim Basin consists of six populations (XLT, XKC, XBC, XKP, XSF and XWQ), the Hami Basin includes populations XQJ and XMG; the western Yumen of Gansu Province (grey line circled in the top part of Fig. 1) harbours populations GLY, GYM, GSB and GAS; and the easternmost region (in the easternmost of the distribution range of *G. przewalskii*) comprises the remaining populations GJT, IWH, NZW and GJY. To test the spatial genetic structure of cpDNA haplotypes, a spatial analysis of molecular variance (SAMOVA, Dupanloup *et al*. 2002) was used to define groups of populations (K) that are geographically homogeneous and genetically differentiated from each other. The analysis was run for K = 2–15, starting from 100 random initial conditions for each run. Finally, the number of groups that maximises the proportion of total genetic variance due to differences among groups of populations (F_{CT}) was retained as the best grouping of populations.

An analysis of molecular variance (AMOVA) (Excoffier *et al*. 1992), and calculation of F_{st} using pairwise distances, were performed to partition variation within and among the defined regions and population groups (identified by SAMOVA) by using ARLEQUIN. Levels of significance were tested by a non-parametric procedure with 1000 permutations. Parameters of population diversity, i.e. average gene diversity within populations (h_S), total genetic diversity across all populations (h_T) and the two parameters of population differentiation (G_{ST}, N_{ST}), were calculated, using PERMUT with 1000 permutations (Pons and Petit 1996). Whereas G_{ST} considers haplotype frequencies, N_{ST} takes into account differences between haplotypes. A significantly higher value of N_{ST} than G_{ST} usually indicates the presence of a phylogeographic structure (Pons and Petit 1996). The *U*-statistic was used to evaluate the significance of difference between N_{ST} and G_{ST} by a permutation test with 1000 permutations.

To test whether *G. przewalskii* underwent recent range expansion, we plotted the mismatch distribution as the observed number of differences between pairs of haplotypes. Unimodal distributions tend to indicate population expansion, whereas more ragged distributions indicate that the population is in stable equilibrium. The sum of squared deviations (SSD) between observed and expected mismatch distributions was used to estimate stepwise expansion models with 1000 parametric bootstrap replicates (Schneider and Excoffier 1999) using ARLEQUIN. A significant SSD ($P \leq 0.05$) was taken as evidence for departure from a model of population expansion. Moreover, Tajima's *D* (Tajima 1989) and Fu's F_S (Fu 1997) were also calculated using ARLEQUIN to test for evidence of range expansions. A significant value for *D* may be due to factors such as population expansion or bottlenecks (Tajima 1996). A significantly large negative value for F_S may be due to population expansion (Fu 1997). Demographic analyses were performed separately for populations in the Tarim Basin and the western Yumen of Gansu Province, given that some of them indicated low genetic diversity or were found to be monomorphic.

Genealogical relationships among all haplotypes were estimated, using the median-joining method, implemented in the Network 4600 program (Bandelt *et al*. 1999). In addition, a neighbour-joining (NJ) tree based on Kimura's (Kimura 1980) two-parameter distance was constructed using the MEGA 4.0 software (Tamura *et al*. 2007). In the NJ analysis, all indels were treated as single mutation events and coded as substitutions (A or T). To evaluate clade support, 1000 replicates of bootstrap analysis (Felsenstein 1985) were performed using fast heuristic search and TBR branch swapping.

Results

Sequence variation

The total length of the three cpDNA regions was 307 bp for *psb*A–*trn*H, 787 bp for *ycf*6–*psb*M and 798 bp for *rpl*32–*trn*L (UAG). Thirty-three substitutions were detected, in positions 81, 110, 111, 115, 126, 127, 130, 131, 135, 139, 140, 143, 178, 221, 471, 846–850, 852, 853, 1258, 1259, 1270–1273, 1294, 1298 and 1660–1662. Sixteen indels (in positions 68–72, 109, 116–123, 170, 186–193, 389–343, 654–659, 742–747, 757–769, 770–782, 853, 1353–1357, 1396, 1398, 1472–1476 and 1701–1723) were coded as single events. In total, 31 haplotypes were distinguished from all individuals analysed, and 10 of these occurred in only one individual (Appendix 1, Fig. 1). All defined haplotype sequences have been deposited in GeneBank databases under Accession numbers (JN887249–JN887259) for *psb*A–*trn*H, (JN887239–JN887248) for *ycf*6–*psb*M and (JN887260–JN887265) for *rpl*32–*trn*L (UAG).

Haplotype distribution

Distribution of the 31 haplotypes was not random but showed strong geographic patterns (Fig. 1, Table 3). In all, 21 haplotypes (11–31) were found in eight populations of the north-western Tarim and Hami Basins, 14 of which (Haplotypes 18–31) occurred in the former. The remaining 10 haplotypes (1–10) were distributed over the other eight populations, of the western Yumen of Gansu Province, Inner Mongolia and Ningxia. Haplotypes 4, 8 and 25 were found in two populations each (Haplotype 4: NZW and GJT; Haplotype 8: GYM and GAS; Haplotype 25: XSF and XWQ). The most widespread haplotype, 18, was carried by 21% of individuals, and was found in XLT, XBC, XKC, XKP and XSF in the north-western Tarim Basin. The next most common, Haplotype 7, was carried by 20% of individuals, and was detected in GLY, GYM, GSB and GAS in the western Yumen of Gansu Province.

Each of the remaining 26 haplotypes was unique to a particular population (Fig. 1, Table 3). Haplotypes 20–31 were all private except for Haplotype 25, and were found in XKP, XSF and XWQ. Haplotypes 11–17 were found in XMG and XQJ,

Table 3. The distribution of 31 chloroplast DNA haplotypes in populations of *Gymnocarpos przewalskii*

Numbers from 1 to 31 correspond to each of the 31 cpDNA haplotypes labelled in Fig. 1. Numbers within populations indicate the number of individuals with that haplotype. Private haplotypes in each population, and populations that do not have private haplotypes, are shown in bold

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
IWH	**4**	**6**																													
NZW			**4**	6																											
GJY					**4**	**6**																									
GJT				10																											
GYM							9	1																							
GSB							10																								
GAS							7	3																							
GLY							6		**3**	**1**																					
XMG											**2**	**4**	**3**	**1**																	
XQJ															**8**	**1**	**1**														
XLT																		10													
XKC																		9	**1**												
XBC																		10													
XKP																		2		**2**	**1**	**5**									
XSF																		1					**1**	**1**	4	**2**	**1**				
XWQ																									1			**2**	**2**	**4**	**1**

and Haplotypes 1–2, 3, 5–6 and 9–10 occurred in IWH, NZW, GJY and GLY, respectively.

In contrast, populations in the north-eastern Tarim Basin (XLT, XKC and XBC) were mainly monomorphic for Haplotype 18, and two of those in the northern Gansu region (GSB and GJT) were monomorphic, for Haplotypes 7 and 4, respectively.

Genetic diversity and population structure

High levels of total genetic diversity across all populations ($h_T = 0.930$, s.e. = 0.0351), and average gene diversity within populations ($h_S = 0.425$, s.e. = 0.0764) were revealed for *G. przewalskii*. Differing levels of genetic variation were detected among the populations (Table 1). Generally, elevated diversities were found in IWH, GJY, GLY, XMG, XKP, XSF and XWQ. From these seven populations, a total of 22 private haplotypes were identified, and the estimated haplotype diversities (H_d) ranged from 0.5333 (IWH and GJY) to 0.8889 (XSF). The estimated mean number of pairwise differences (Π) ranged between 9.6667 (GLY) and 24.0000 (IWH and GJY), and nucleotide diversity (Π_n) ranged from 0.0054 (GLY) to 0.0132 (IWH). In contrast, medium or low levels of genetic diversity were identified in XQJ, NZW and XKC, but these populations possessed three (100%), one (50%) and one (50%) unique haplotypes, respectively. No cpDNA variation was found in populations GJT, GSB, XLT or XBC.

The *U*-test showed that N_{ST} (0.752, s.e. = 0.055) was significantly larger than G_{ST} (0.547, s.e. = 0.072) ($U = 1.36$, $P < 0.01$), indicating significant phylogeographic structure across the species' range. AMOVA results supported divergence of the defined four regions, with ~63% of the variation attributed to this pattern of differentiation. Spatial genetic analysis of cpDNA haplotypes using SAMOVA indicated that F_{CT} increased to a maximal value of 0.7835 when K (the number of groups) was raised from $K = 2$ to $K = 12$. The grouping pattern of populations corresponding to $K = 12$ was (1) NZW, (2) IWH, (3) GJY, (4) GJT, (5) GYM, GSB, GAS, (6) GLY, (7) XMG, (8) XQJ, (9) XLT, XKC, XBC, (10) XKP, (11) XSF and (12) XWQ (Table 4). AMOVA showed that 77.87% of total genetic variation was partitioned among the 12 groups of populations, and only 23.87% occurred within groups.

Table 4. Analyses of molecular variance (AMOVA) of chloroplast DNA sequences from *Gymnocarpos przewalskii* populations

Twelve groups: (1) NZW, (2) GJY, (3) GJT, (4) GLY, (5) IWH, (6) GYM, GSB, GAS, (7) XMG, (8) XQJ, (9) XLT, XKC, XBC, (10) XKP, (11) XSF and (12) XWQ. *$P < 0.01$

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among four regions	3	1514.583	12.25976 Va	63.00*
Among populations with regions	12	418.842	3.07799 Vb	15.81
Within populations	144	593.800	4.12361 Vc	21.19*
Total	159	2527.225	19.46136	
Among twelve groups	11	1928.958	13.45035 Va	77.87*
Among populations within groups	4	4.467	0.30069 Vb	–1.74
Within populations	144	593.800	4.12361 Vc	23.87*
Total	159	2527.225	17.27327	

Phylogeographic analysis

The haplotype network illustrates the complex relationships between the 31 haplotypes (Fig. 2). Two large groups of haplotypes, representing those occurring in the Hami Basin, western Yumen of Gansu Province, Inner Mongolia and Ningxia are shown on the left (the eastern group), and haplotypes mainly occurring in the Tarim Basin are shown on the right (the western group). The most common haplotypes, namely Haplotypes 7 and 18, were respectively widespread in these two groups. Additionally, haplotypes of the eastern group were found to be subdivided into four subgroups. Except for Haplotype 10 (shown as white in Fig. 2), the remaining represent the following three geographical areas: the easternmost region,

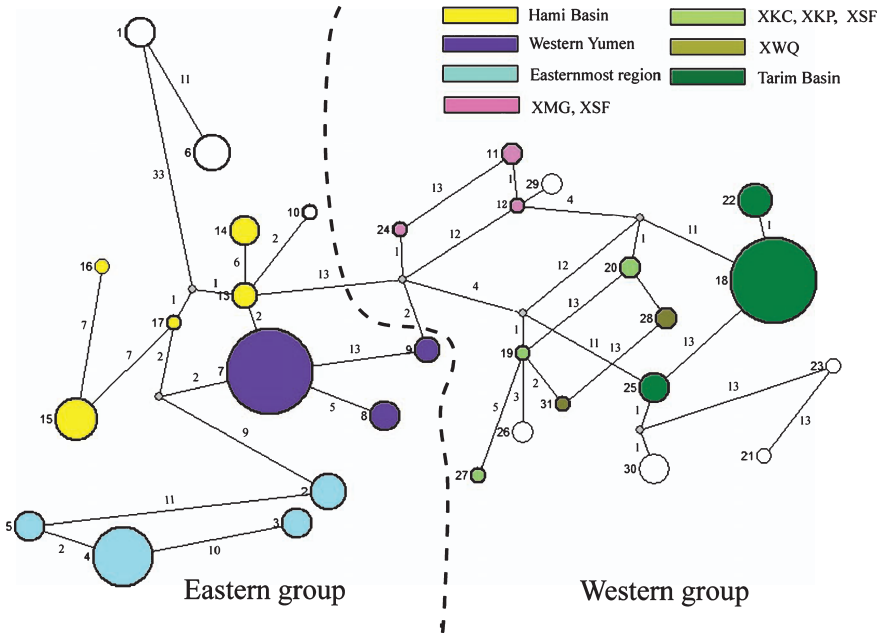

Fig. 2. Haplotype network of *Gymnocarpos przewalskii* individuals collected from 16 different populations. The grey circles indicate the missing or inferred haplotypes. The circle areas are proportional to haplotype frequencies. Branch lengths are roughly proportional to the number of mutation steps between haplotypes and nodes; the true number of steps is shown near the corresponding branch section. The groups of haplotypes that are consistent with the connections of neighbour-joining (NJ) tree (Fig. 3) are indicated with different colours, as illustrated in the respective corresponding legends.

Haplotypes 1 and 6 (pale blue and white), Hami Basin (yellow) and the western Yumen of Gansu Province (purple). In the western group, the distributions of haplotypes presented a complex pattern. Haplotypes 11 and 12 from XMG were connected with Haplotype 24 of XSF and Haplotype 29 of XWQ. In addition, Haplotypes 23–27 were all from the same population, XSF, but did not cluster together in the network. Haplotypes 20–22 of XKP and Haplotypes 28–31 of XWQ are also similar cases.

The neighbour-joining (NJ) tree identified five clades that received over 60% bootstrap support, but these did not include the small clade consisting of Haplotypes 28, 29 and 31 (Fig. 3). The first two clades included haplotypes mainly from the Tarim Basin, except for Haplotype 10 of GLY. The Hami Basin clade comprised Haplotypes 13–17 of XMG and XQJ, and also Haplotype 1 of IWH and Haplotype 24 of XSF. However, the western Yumen and easternmost clades were composed of haplotypes that were all from their respective regions. In general, the identified large-scale geographic structuring of haplotypes in the NJ tree is similar to that displayed in the network (Figs 2, 3).

Demographic analyses

The observed mismatch distributions for cpDNA haplotypes, calculated for the populations in the Tarim Basin and western Yumen of Gansu Province (Fig. 4), were not unimodal, and thus differed strongly from the prediction under a model of sudden range expansion. This difference was also supported by a significant SSD statistic (Tarim Basin: SSD = 0.055, $P = 0.041$; western Yumen: SSD = 0.191, $P = 0.018$). The lack of population expansion was further supported by non-significant D and F_s values (Tarim Basin: $D = 0.319$, $P = 0.770$; $F_s = 3.791$, $P = 0.977$; western Yumen: $D = -0.490$, $P = 0.751$; $F_s = 3.904$, $P = 0.950$). Thus, no evidence is provided for recent demographic expansions in *G. przewalskii*.

Discussion

Genetic variation in G. przewalskii

Of the 31 different haplotypes identified from the 16 populations, 26 were private and found in 10 of the populations. The levels of total genetic diversity ($h_T = 0.903$) across all populations, and average within-population diversity ($h_S = 0.425$) were very elevated, compared with, for example, variation in three cpDNA regions in the small shrub *Juniperus sabina* from northern China ($h_T = 0.577$, $h_S = 0.043$; Guo *et al*. 2010). The high cpDNA diversity found in *G. przewalskii* could be the result of its ancient presence in this area (Fu 1992), allowing the accumulation of a significant number of mutations (Yuan *et al*. 2008; Falchi *et al*. 2009). Additionally, the diversity in habitats, resulting from several episodes of rapid aridification since the Quaternary, and greater geological and topological variation occurring in north-western China (Guo *et al*. 1999), may have promoted variability by selection of new mutations.

Both AMOVA and SAMOVA analyses showed significant genetic differentiation among the sampled populations. These differences appear to be strongly associated with pollination and seed dispersal. Gene flow between populations via pollen would be limited by the small quantities of pollen and a low frequency of insect pollination (Wang *et al*. 2009*b*). Also, the seeds are clumped and retained together until germination, and their dispersal from the persistent calyx is constrained by gravity

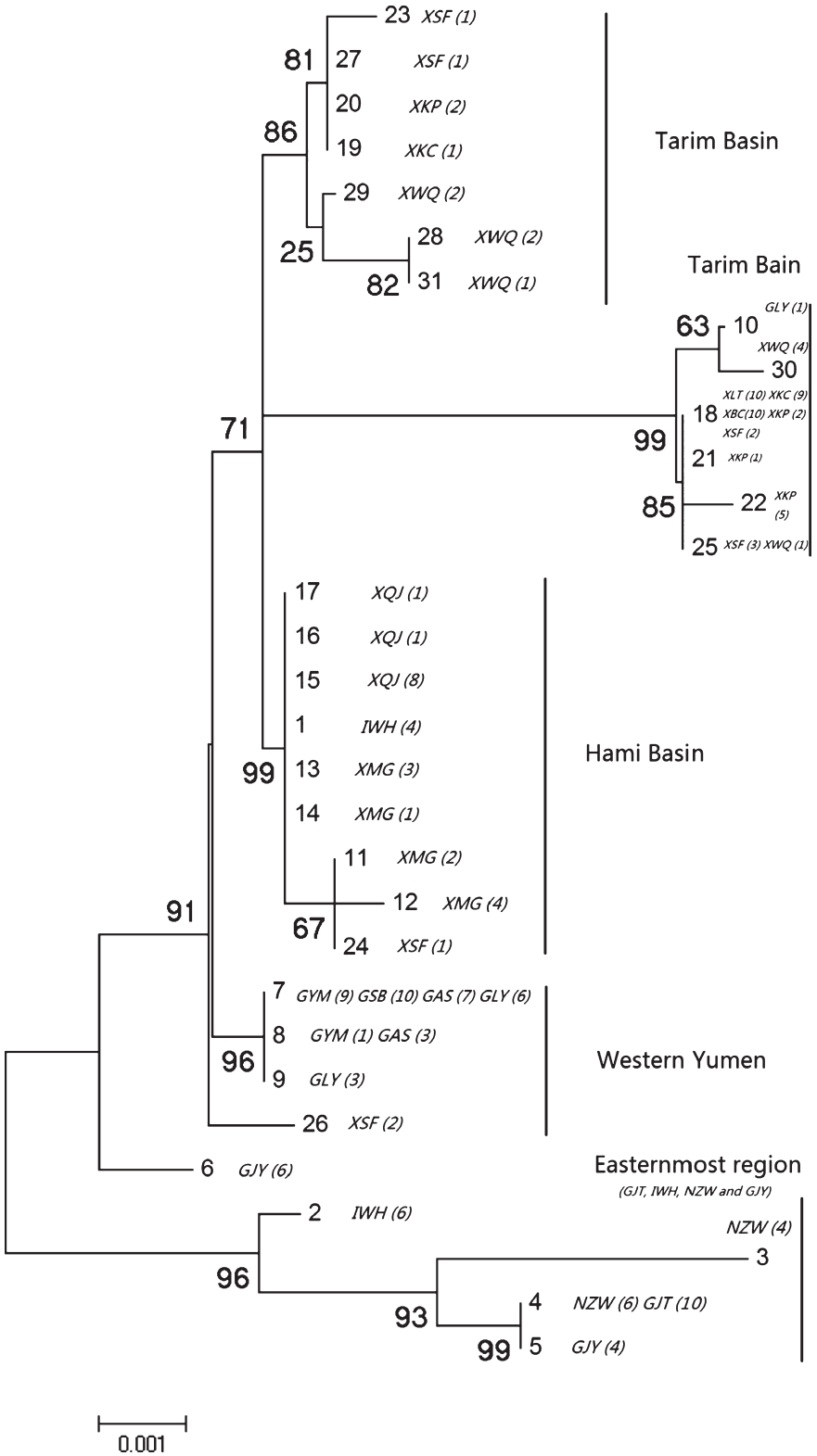

Fig. 3. Neighbour-joining (NJ) tree of *Gymnocarpos przewalskii*, based on cpDNA variation of the *psb*A–*trn*H, *ycf*6–*psb*M and *rpl*32–*trn*L (UAG) spacers. Numbers at nodes show bootstrap values generated from the bootstrap of the NJ tree (≥60%, 1000 replicates). The characters in italics following the haplotype number and the numbers in brackets show the population name (i.e. the population codes listed in Table 1) and number of individuals in the corresponding haplotype, respectively. The black bars on the right indicate the corresponding clades, representing the following four regions of haplotypes: Tarim Basin, Hami Basin, western Yumen and the easternmost region (the easternmost distribution range includes the GJT, IWH, NZW and GJY populations).

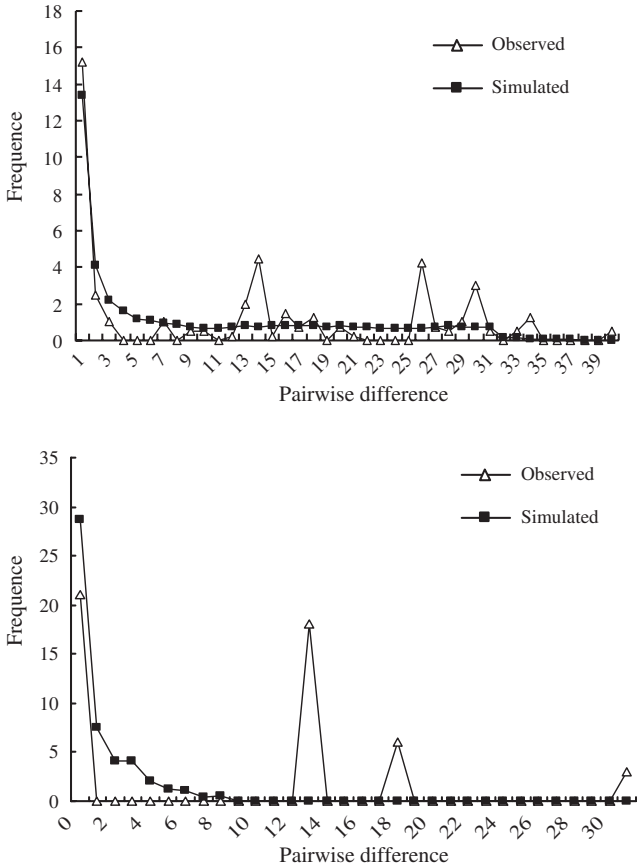

Fig. 4. Mismatch distributions for the Tarim Basin (above) and western Yumen (below) were identified by SAMOVA. The line with triangles shows observed values, whereas the line with solid squares represents expected values under a model of sudden (stepwise) expansion.

(Van der Pijl 1969), likely resulting in most seed dispersal being confined to short distances.

Allopatric divergence

The strong phylogeographic structure ($N_{ST} > G_{ST}$, $P < 0.01$) found in *G. przewalskii* reflects the fact that separate lineages occupy the different geographic regions. Four distinct haplotype groupings can be distinguished, and these are consistent with their respective geographical distributions in the Tarim Basin, Hami Basin, the western Yumen of Gansu Province and the easternmost region (i.e. regions of GJT, IWH, NZW and GJY). Results from AMOVA suggested that over 60% of the observed variation was due to differences among these regions. Moreover, SAMOVA analysis showed that each region was subdivided into two or four groups of populations, so that a total of 12 population groups was found to be optimal (red lines circled in Fig. 1). AMOVA also supported divergence of the 12 groups, with ~77% of the total variation attributed to this pattern of differentiation. The subdivision could be related to specific geologic histories within each of the four regions. Our results suggest that regional genetic differentiation of *G. przewalskii* has resulted mainly from geographic isolation posed by the mountains and large deserts that developed in north-western China, and from range contraction and population fragmentation induced by climate oscillations.

The eastern extension of the Tian Shan Mountains separates the enclosed intermontane Hami Basin from the inland Tarim Basin. This basin–mountain geographic pattern would create a large impediment to gene flow and the spread of the species, resulting in a high degree of population isolation in these basins. Likewise, the HeLan Mountains appear to have effectively prevented a north–south mediated gene flow between IWH and NZW, and the Wang Mountains in the north of the Jingyuan region of southern Gansu might have served as a geographical barrier between GJY and NZW, blocking genetic exchanges between the isolated populations. In addition, the aridity of north-western China as a whole has been strengthened since the Quaternary, when the deserts rapidly expanded on a large scale (Zhang *et al*. 2000, Fang *et al.* 2002). The arid conditions have caused habitat fragmentation and population isolation of many plant species in this area (e.g. Chang *et al*. 2004; Chen *et al*. 2005). For *G. przewalskii*, the aridity coupled with the great expansion of Badain Jaran–Tengger Desert during the middle Pleistocene (Yang *et al*. 2006), and the gradual westward expansion of Taklimakan Desert since the early Pleistocene (Mu 1994), may have together affected the current allopatric divergence between the Tarim Basin, Hami Basin, western Yumen of Gansu Province and the easternmost region (Ma *et al*. 2010).

Recolonisation and glacial refugia

Low levels of genetic variation in *G. przewalskii* are found in the northern Tarim Basin (XBC, XKC and XLT), western Yumen (GAS, GSB and GYM), and in sand dunes of the Jinta region (GJT) of Gansu Province (Table 1). These areas are mostly dominated by Haplotypes 18, 7 or 4. This is consistent with previous reports, that post-glacially colonised regions are expected to have reduced levels of genetic variation, with large geographic areas that mainly harbour a single haplotype (Hewitt 2000). However, no significant evidence of population expansions was found in *G. przewalskii*. As indicated by Printzen *et al*. (2003), large-scale intraspecific disjunctions in many species could be explained alternatively by range fragmentation and long-distance dispersal and colonisation. In our case, *G. przewalskii* is a clonal shrub. Vegetative reproduction of this species is by sand-covered branches, which could possibly have allowed long-distance colonisation owing to the action of animals. The several expansions of the Taklimakan and the Badain Jaran–Tengger Deserts may have facilitated the demographic fluctuations of *G. przewalskii*, that is, population bottlenecks followed by subsequent post-glacial recolonisation, which would have elevated genetic drift, and lead to a loss of genetic variation (Barker *et al*. 2009). In the haplotype network, interior positions coupled with high-frequency occurrence indicate that Haplotypes 7 and 18 were ancestral (Fig. 2). Except for Haplotype 8, each of the haplotypes at tip positions was unique to a particular population. According to coalescent theory (Crandall and Templeton 1993), Haplotypes 7 and 18 might represent the relics retained in large populations in different regions through time, whereas haplotypes situated at tips in the network would be more likely to have arisen during colonisation.

No haplotype was found to be shared among the four main regions, or between populations within the Hami Basin and those of the easternmost region. Instead, several sets of unique haplotypes were found in these groups of populations. According to Petit *et al*. (2003), high genetic divergence and uniqueness in plant populations indicate that the associated areas may have served as sites of glacial refugia. Phylogeographic studies on several other rare and/or endangered species in China have revealed several refugia in southern China, and detected evidence of colonisation following glacial periods (e.g. Shen *et al*. 2005; Wang and Ge 2006; Wang *et al*. 2009*a*). However, relatively few glacial refugia have been detected by these studies. One case is the western area of Tarim basin, which was deemed a refugium during the LGM, according to palynological evidence (Fan 1993); this was also indicated as a refugium for the Yarkand hare, a species with a widespread distribution in that basin (Shan *et al*. 2011). The other is the north-western area of Inner Mongolia, which has high endemism (e.g. *Tetraena mongolica*, *Ammopiptanthus mongolicus* and *Potaninia mongolica*), and has been identified as a possible refugium during the LGM (Zhao 1997). In our study, unique haplotypes and high levels of genetic variation were found near the western Tarim Basin (XWQ, XSF and XKP regions), the Hami Basin (XMG region), the Liu Yuan region in the western Gansu (GLY) and our easternmost region (NZW, IWH and GJY regions), which suggests that these areas must have played key roles as glacial refugia for *G. przewalskii*. In this regard, therefore, this species may have survived the LGM in at least four independent refugia that have given rise to the distinctive cpDNA lineages that characterise the different components of the species' present-day distribution.

Conservation implications

In the present study, the genealogical relationships between haplotypes of *G. przewalskii*, as well as their geographic distribution across the species range were elucidated (Figs 1–3). The identification of historically isolated, genetically divergent lineages is important for the development of plans, to ensure conservation of the species. Our findings indicate that significant genetic differentiation exists among four main regions and among the 12 population groups identified by SAMOVA (Table 4). To capture a considerable amount of the overall genetic variation of *G. przewalskii*, all of the identified haplotype lineages (31 haplotypes) should be conserved, although with different conservation priorities. Specifically, refugial populations XWQ, XSF and XMG, possessing four unique haplotypes each, should be of first priority for conservation, followed by XKP, which harboured three unique haplotypes, and then GLY, IWH and GJY, which harboured two private haplotypes each. These populations have the greatest possibility for persevering genetic variation under future climate changes and are critical for maintaining evolutionary potential. Additionally, the XQJ, NZW and XKC populations have three or one unique haplotypes, respectively; and GYM and GAS share Haplotype 8, which was not found elsewhere. All are important to conserve, to maintain the genetic heterogeneity of the species. The monomorphic populations XLT, XBC, GSB and GJT should be given lower priority for conservation.

For *ex situ* conservation, populations at the refugial sites mentioned above could serve as seed sources. Seed collections from these populations should follow standard and well recognised guidelines (Zawko *et al*. 2001; Chen *et al*. 2009). For example, they should include as many plants as possible to obtain quantitative genetic variation, and plants should be collected from as distant localities as possible, to avoid inbreeding depression. Such approaches are necessary to maximise the probability of successful reintroduction of *G. przewalskii* if it becomes extinct in any areas in the wild. For *in situ* conservation, management plans would be expected to focus on the maintenance of effective population sizes and reduction of human disturbance.

Conclusion

Strong phylogeographic patterns were detected in *G. przewalskii*, on the basis of the cpDNA spacers *psb*A–*trn*H, *ycf*6–*psb*M, and *rpl*32–*trn*L (UAG). Glacial refugia were inferred along the western Tarim Basin, in the Hami Basin, in the Liuyuan region of western Gansu and around the easternmost region (the Wulate Rear Banner region in Inner Mongolia, Jingyuan region in southern Gansu and the Zhongwei region in western Ningxia). Population bottlenecks followed by subsequent post-glacial recolonisation were identified in the northern Tarim Basin, and the western Yumen and Jinta regions of northern Gansu. Conservation strategies for the identified haplotype lineages were proposed. The phylogeography of *G. przewalskii* presented in the current study provides basic information on how plants in the deserts of north-western China have responded to Quaternary climatic oscillations. However, more similar studies are needed before general conclusions can be drawn.

Acknowledgements

We thank Hong-xiang Zhang, Hong-hu Meng and Xiao-jun Shi for their assistance with field survey and sample collection. Funding was provided by CAS Important Direction for Knowledge Innovation Project (No. KZCX2-EW-305), and Xinjiang Institute of Ecology and Geogeraphy, CAS.

References

Avise JC (2000) 'Phylogeography: the history and formation of species.' (Harvard University Press: Cambridge, MA)

Bandelt HJ, Forster P, Röhl A (1999) Median joining networks for inferring intraespecific phylogenies. *Molecular Biology and Evolution* **16**, 37–48.

Barker JS, Frydenberg J, González J, Davies HI, Ruiz A, Sørensen JG, Loeschcke V (2009) Bottlenecks, population differentiation and apparent selection at microsatellite loci in Australian *Drosophila buzzatii*. *Heredity* **102**, 389–401. doi:10.1038/hdy.2008.127

Chang ZY, Xu LR, Wu ZH (2004) Comparative morphology, ecology and geographical distribution of *Calophaca sinica* and *C. soongorica*. *Acta Botanica Boreali Occidentalia Sinica* **24**, 2312–2320. [in Chinese, with English abstract]

Chai YQ, Chao ZZ, Cai ZS (2010) A study on species diversity of a *Gymnocarpos przewalskii* community in the Subei desert. *Acta Prataculturae Sinica* **19**, 21–27. [in Chinese, with English abstract]

Chen GQ, Huang HW, Ge XJ (2005) Allozyme diversity and population differentiation in an endangered plant, *Ammopiptanthus nanus* (Leguminosae). *Journal of Wuhan Botanical Research* **23**, 131–137. [in Chinese, with English abstract]

Chen GQ, Huang HW, Crawford DJ, Pan BR, Ge XJ (2009) Mating system and genetic diversity of a rare desert legume *Ammopiptanthus nanus* (Leguminosae). *Journal of Systematics and Evolution* **47**, 57–66. doi:10.1111/j.1759-6831.2009.00003.x

Comes HP, Kadereit JW (1998) The effect of Quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science* **3**, 432–438.

Crandall KA, Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* **134**, 959–969.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**, 11–15.

Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* **11**, 2571–2581. doi:10.1046/j.1365-294X.2002.01650.x

Dutech C, Maggia L, Joly HI (2000) Chloroplast diversity in *Vouacapoua Americana* (Caesalpiniaceae), a neotropical forest tree. *Molecular Ecology* **9**, 1427–1432. doi:10.1046/j.1365-294x.2000.01027.x

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes, applications to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* **1**, 47–50.

Falchi A, Paolini J, Desjobert JM, Melis A, Costa J, Varesi L (2009) Phylogeography of *Cistus creticus* L. on Corsica and Sardinia inferred by the *trn*L–F and *pl*32–*trn*L sequences of cpDNA. *Molecular Phylogenetics and Evolution* **52**, 538–543. doi:10.1016/j.ympev.2009.04.002

Fang XM, Lu LQ, Yang SL, Li JJ, An ZS, Jiang P, Chen XL (2002) Loess in Kunlun Mountains and its implications on desert development and Tibetan Plateau uplift in west China. *Science China Earth Sciences* **45**, 289–299. [Science in China Series D]. doi:10.1360/02yd9031

Fan ZL (1993) A study on the formation and evolution of oases in Tarim Basin. *Acta Geographica Sinica* **5**, 421–427. [in Chinese, with English abstract]

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791. doi:10.2307/2408678

Fu LG (1992) 'China plant red data book.' (Science Press: Beijing) [in Chinese].

Fu XY (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925.

Ge XJ, Yu Y, Yuan YM, Huang HW, Yan C (2005) Genetic diversity and geographic differentiation in endangered *Ammopiptanthus* (Leguminosae) populations in desert regions of northwest China as revealed by ISSR analysis. *Annals of Botany* **95**, 843–851. doi:10.1093/aob/mci089

Ge XJ, Hwang CC, Liu ZH, Huang CC, Huang WH, Huang KH, Wang WK, Chiang TY (2011) Conservation genetics and phylogeography of endangered and endemic shrub *Tetraena mongolica* (Zygophyllaceae) in Inner Mongolia, China. *BMC Genetics* **12**, 1–12. doi:10.1186/1471-2156-12-1

Guo ZT, Peng SZ, Hao QZ, Chen XH, Liu TS (1999) Late Tertiary development of aridification in northwestern China: link with the arctic ice-sheet formation and Tibetan uplifts. *Quaternary Sciences* **6**, 556–566. [in Chinese, with English abstract]

Guo YP, Zhang R, Chen CY, Zhou DW, Liu JQ (2010) Allopatric divergence and regional range expansion of *Juniperus sabina* in China. *Journal of Systematics and Evolution* **48**, 153–160. doi:10.1111/j.1759-6831.2010.00073.x

Hamilton M (1999) Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology* **8**, 521–523.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913. doi:10.1038/35016000

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**, 111–120. doi:10.1007/BF01731581

Li N, Fu L (1997) Notes on gymnosperms I. Taxonomic treatment of some Chinese conifers. *Novon* **7**, 261–264. doi:10.2307/3391941

Li WC (1998) 'The Chinese Quaternary vegetation and environment.' (Science Press: Beijing) [in Chinese].

Ma SM, Zhang ML, Zhang HX, Meng HH, Chen X (2010) Predicting potential geographical distributions and patterns of the relic plant *Gymnocarpos przewalskii* using maximum entropy and genetic algorithm for rule-set prediction. *Chinese Journal of Plant Ecology* **34**, 1327–1335. [in Chinese, with English abstract]

Mu GJ (1994) On the age and evolution of the Taklimakan desert. *Arid Land Geography* **17**, 1–9. [in Chinese, with English abstract]

Nei M (1987) 'Molecular evolutionary genetics.' (Columbia University Press: New York)

Oxelman B, Ahlgren B, Thulin M (2002) Circumscription and phylogenetic relationships of *Gymnocarpos* (Caryophyllaceae–Paronychioideae). *Edinburgh Journal of Botany* **59**, 221–237. doi:10.1017/S0960428602000045

Petit RJ, Aguinagalde I, de Beaulieu JL, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Muller-Starck GM, Demesure-Musch B, Palme A, Martin P, Rendell S, Vendramin GG (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**, 1563–1565. doi:10.1126/science.1083264

Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* **144**, 1237–1245.

Printzen C, Ekman S, Tonsberg T (2003) Phylogeography of *Cavernularia hultenii*: evidence of slow genetic drift in a widely disjunct lichen. *Molecular Ecology* **12**, 1473–1486. doi:10.1046/j.1365-294X.2003.01812.x

Raspé O, Saumitou-Laprade P, Cuguen J, Jacquemart AL (2000) Chloroplast DNA haplotype variation and population differentiation in *Sorbus aucuparia* L. (Rosaceae: Maloideae). *Molecular Ecology* **9**, 1113–1122. doi:10.1046/j.1365-294x.2000.00977.x

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496–2497. doi:10.1093/bioinformatics/btg359

Sang T, Crawford DJ, Stuessy TF (1997) Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* **84**, 1120–1136. doi:10.2307/2446155

Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* **152**, 1079–1089.

Shackleton NJ, Opdyke ND (1973) Oxygen isotope and palaeomagnetic stratigraphy of equatorial Pacific core v28–238 isotope temperatures and ice volumes on a 105 year and 106 year scale. *Quaternary Research* **3**, 39–55. doi:10.1016/0033-5894(73)90052-5

Shan WJ, Liu J, Yu L, Robert WM, Mahmut H, Zhang YP (2011) Genetic consequences of postglacial colonization by the endemic Yarkand hare (*Lepus yarkandensis*) of the arid Tarim Basin. *Chinese Science Bulletin* **56**, 1370–1382. doi:10.1007/s11434-011-4460-9

Shaw J, Lickey E, Schilling E, Small R (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms, the tortoise and the hare III. *American Journal of Botany* **94**, 275–288. doi:10.3732/ajb.94.3.275

Shaw J, Small RL (2005) Chloroplast DNA phylogeny and phylogeography of the North American plums (*Prunus* subgenus *Prunus* section *Prunocerasus*, Rosaceae). *American Journal of Botany* **92**, 2011–2030.

Shen L, Chen XY, Li YY (2002) Glacial refugia and postglacial recolonization patterns of organisms. *Acta Ecologica Sinica* **22**, 1983–1990. [in Chinese, with English abstract]

Shen L, Chen XY, Zhang X, Li YY, Fu CX, Qiu YX (2005) Genetic variation of *Ginkgo biloba* L. (Ginkgoaceae) based on cpDNA PCR–RFLPs: inference of glacial refugia. *Heredity* **94**, 396–401. doi:10.1038/sj.hdy.6800616

Tajima F (1983) Evolutionary relationships of DNA sequences in finite populations. *Genetics* **105**, 437–460.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.

Tajima F (1996) The amount of DNA polymorphism maintained in a finite population when the neutral mutation rate varies among sites. *Genetics* **143**, 1457–1465.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596–1599. doi:10.1093/molbev/msm092

Thompson JDTJ, Gibson PF, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**, 4876–4882. doi:10.1093/nar/25.24.4876

Van der Pijl L (1969) 'Principles of dispersal in higher plants.' (Springer Verlag Press: Berlin)

Vidal-Russell R, Souto CP, Premoli AC (2011) Multiple Pleistocene refugia in the widespread Patagonian tree *Embothrium coccineum* (Proteaceae). *Australian Journal of Botany* **59**, 299–314.

Wang GH (2005) The western Ordos plateau as a biodiversity center of relic shrubs in arid areas of China. *Biodiversity and Conservation* **14**, 3187–3200. doi:10.1007/s10531-004-0386-8

Wang HW, Ge S (2006) Phylogeography of the endangered *Cathaya Argyrophylla* (Pinaceae) inferred from sequence variation of mitochondrial and nuclear DNA. *Molecular Ecology* **15**, 4109–4122. doi:10.1111/j.1365-294X.2006.03086.x

Wang LY, Abbott RJ, Zheng W, Chen P, Wang YJ, Liu JQ (2009*a*) History and evolution of alpine plants endemic to the Qinghai–Tibetan Plateau: *Aconitum gymnandrum* (Ranunculaceae). *Molecular Ecology* **18**, 709–721. doi:10.1111/j.1365-294X.2008.04055.x

Wang ZB, Gao QX, Sun JZ, Ma QL (2009*b*) Study on biological characteristics of rare endangered plant *Gymnocarpos Przewalskii. Resource Development & Market* **25**, 481–509. [in Chinese, with English abstract]

Yang D, Fang XM, Dong GR, Peng ZC, Li JJ (2006) Aeolian deposit evidence for formation and evolution of the Tengger Desert in the north of China since early Pleistocene. *Marine Geology and Quaternary Geology* **26**, 93–100. [in Chinese, with English abstract]

Yang Y, Liu M (2002) Cenozoic deformation of the Tarim plate and the implications for mountain building in the Tibetan Plateau and the Tian Shan. *Tectonics* **21**, 1059–1076. doi:10.1029/2001TC001300

Yuan QJ, Zhang ZY, Peng H, Ge S (2008) Chloroplast phylogeography of *Dipentodon* (Dipentodontaceae) in southwest China and northern Vietnam. *Molecular Ecology* **17**, 1054–1065. doi:10.1111/j.1365-294X.2007.03628.x

Zawko G, Krauss SL, Dixon KW, Sivasithamparam K (2001) Conservation genetics of the rare and endangered *Leucopogon obtectus* (Ericaceae). *Molecular Ecology* **10**, 2389–2396. doi:10.1046/j.0962-1083.2001.01378.x

Zhang DF, Fengquan L, Jianmin B (2000) Eco-environmental effects of the Qinghai–Tibet Plateau uplift during the Quaternary in China. *Environmental Geology* **39**, 1352–1358. doi:10.1007/s002540000174

Zhao YZ (1997) Endemic genera and their basic characteristics of the Mongolian Plateau plants. *Acta Scientiarum Naturalium Universitatis NeiMongol* **28**, 547–552. [in Chinese, with English abstract]

Appendix 1. Variable nucleotide sites in three chloroplast DNA regions in 31 haplotypes of *Gymnocarpos przewalskii*

◆, GTTT; ■, CGGTTTC; ▼, TTTGATA; ●, GTTTT; ★, ATAGAA; ▲, ATAGTT; ⊕, TTTATTATAGTTT; #, TTTATTATAGTTT; ⊀, TATTT; ∧, ACCAA; ∞, TTAATGTTCTATTTATGTTAAAA. Open symbols, i.e. ◇, □, ▽, ○, ☆ and △, and ε, ⊿, ⊁, ∨ and ∝ represent the indels format text, with the length >1 bp. '–' represenst indels, with the length equal to 1 bp

*psb*A–*trn*H																					*ycf*6–*psb*M											*rpl*32–*trn*L (UAG)															
68	81	109	110	111	115	116	126	127	130	131	135	139	140	143	170	178	186	221	389	471	654	742	757	770	846	847	848	849	850	851	853	1258	1259	1270	1271	1272	1273	1294	1298	1353	1396	1398	1472	1660	1661	1662	1701
◇	G	–	G	G	C	□	A	C	A	A	G	T	T	G	G	T	▽	C	○	A	☆	△	ε	⊿	A	T	A	G	A	T	C	G	A	A	A	A	G	A	A	⊀	A	–	∧	T	A	G	∞
◇	.	–	.	.	.	□	.	.	.	.	.	.	.	.	.	.	▽	.	○	.	☆	△	ε	⊿	.	.	.	.	.	.	.	T	T	C	T	T	T	T	C	⊁	–	.	∨	G	C	T	∝
◇	.	C	A	A	.	□	T	T	G	T	.	C	C	.	–	.	▽	.	○	.	★	△	ε	⊿	T	C	T	A	T	.	.	T	T	C	T	T	T	T	C	⊁	.	A	∨	G	C	T	∝
◇	.	–	.	.	.	□	.	.	.	.	.	.	.	.	.	.	▽	.	○	.	★	△	ε	⊿	T	C	T	A	T	.	.	T	T	C	T	T	T	T	C	⊁	.	A	∨	G	C	T	∝
◇	.	–	.	.	.	□	.	.	.	.	.	.	.	.	.	.	▽	.	○	.	★	△	ε	⊿	T	C	T	A	T	.	.	T	T	C	T	T	T	T	C	⊁	–	.	∨	G	C	T	∝
◇	.	.	.	.	.	□	.	.	.	.	.	.	.	.	.	.	▽	.	○	.	★	△	ε	⊿	T	C	T	A	T	.	.	.	.	.	.	.	.	.	.	⊀	.	.	∧	.	.	.	∞
◇	.	–	.	.	.	□	.	.	.	.	.	.	.	.	.	.	▽	.	○	.	☆	△	ε	⊿	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	C	T	∝
◇	.	–	.	.	.	□	.	.	.	.	.	.	.	.	.	.	▽	.	○	.	☆	▲	ε	⊿	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	C	T	∝
◇	.	–	.	.	.	□	.	.	.	.	.	.	.	.	.	.	▽	.	○	.	☆	△	⊕	⊿	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	C	T	∝
◆	A	C	A	A	A	□	T	T	G	T	.	C	C	.	–	.	▽	.	○	.	☆	△	ε	⊿	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	.	.	∝
◇	.	–	.	.	.	□	.	.	.	.	.	.	.	.	.	.	▽	.	○	.	☆	△	⊕	#	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	.	.	∝
◇	.	.	.	.	.	□	.	.	.	.	.	.	.	.	.	.	▽	T	○	.	☆	△	ε	⊿	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	.	.	∝
◇	.	–	.	.	.	□	.	.	.	.	.	.	.	.	.	.	▽	T	○	.	☆	△	ε	⊿	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	.	.	∝
◇	.	–	.	.	.	□	.	.	.	.	.	.	.	.	.	.	▽	.	○	.	☆	△	⊕	#	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	.	.	∝
◇	.	.	.	.	.	■	.	.	.	.	.	.	.	.	.	.	▽	.	○	.	☆	△	ε	⊿	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	–	.	∨	.	.	.	∝
◇	.	.	.	.	.	■	.	.	.	.	.	.	.	.	.	.	▼	.	○	.	☆	△	ε	⊿	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	–	.	∨	.	.	.	∝
◇	.	–	.	.	.	□	.	.	.	.	.	.	.	.	.	.	▽	.	○	.	☆	△	ε	⊿	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	–	.	∨	.	.	.	∝
◆	.	C	A	A	A	□	T	T	G	T	.	C	C	.	–	.	▽	.	○	.	☆	△	⊕	#	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	.	.	∝
◆	.	–	.	.	.	□	.	.	.	.	T	.	.	.	.	.	▽	.	○	.	☆	△	⊕	⊿	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	.	.	∝
◆	.	–	.	.	.	□	.	.	.	.	T	.	.	.	.	.	▽	.	○	.	☆	△	⊕	#	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	.	.	∝
◆	.	C	A	A	A	□	T	T	G	T	.	C	C	.	–	.	▽	.	○	.	☆	▲	ε	⊿	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	.	.	∝
◆	.	C	A	A	A	□	T	T	G	T	.	C	C	.	–	G	▽	.	○	.	☆	△	⊕	#	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	.	.	∝
◆	.	–	.	.	.	□	A	C	A	A	T	.	.	.	.	.	▽	.	○	.	☆	△	ε	⊿	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	.	.	∝
◇	.	–	.	.	.	□	A	C	A	A	.	.	.	.	.	.	▽	T	○	.	☆	△	⊕	⊿	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	.	.	∝
◆	.	C	A	A	A	□	T	T	G	T	.	C	C	.	–	.	▽	.	○	.	☆	△	⊕	⊿	.	.	.	.	.	–	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	.	.	∝
◆	.	–	.	.	.	□	.	.	.	.	T	.	.	.	.	.	▽	.	○	.	☆	△	⊕	⊿	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	.	.	∝
◆	.	–	.	.	.	□	.	.	.	.	T	.	.	.	.	.	▽	.	●	.	☆	△	⊕	⊿	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	.	.	∝
◆	A	–	.	.	.	□	.	.	.	.	T	.	.	.	.	.	▽	.	○	T	☆	△	⊕	#	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	.	.	∝
◆	.	–	.	.	.	□	.	.	.	.	.	.	.	.	.	.	▽	.	○	T	☆	△	⊕	#	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	.	.	∝
◇	A	C	A	A	A	□	T	T	G	T	.	C	C	.	–	.	▽	.	○	T	☆	△	⊕	⊿	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	.	.	∝
◆	A	–	.	.	.	□	.	.	.	.	T	.	.	.	.	.	▽	.	○	T	☆	△	⊕	⊿	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	⊁	.	A	∨	.	.	.	∝