Conservation genetics and geographic patterns of genetic variation of the endangered officinal herb *Fritillaria pallidiflora*

Zhihao Su, Borong Pan, Stewart C. Sanderson, Xiaolong Jiang and Mingli Zhang

*Fritillaria pallidiflora* is an endangered officinal herb distributed in the Tianshan Mountains of northwestern China. We examined its phylogeography to study evolutionary processes and suggest implications for conservation. Six haplotypes were detected based on three chloroplast non-coding spacers (*psbA-trnH, rps16*, and *trnS–trnG*); genetic variation mainly occurred among populations and SAMOVA groups. This species is distributed in different deep valleys, and we speculate that these fragmented habitats cause gene flow barriers among populations and groups. We also speculate that during glacial periods, extremely low temperatures and aridity caused additional range shrinkage and fragmentation, factors consequently resulting in significant intraspecific differentiation in allopatric regions. For setting a conservation management plan, we identified the Lucaogou region as the most important area, and we designated two ESUs for separate management.

Genetic variation within and among natural populations is crucial for the long-term survival of a species. Accurate information about genetic diversity of threatened species provides fundamental data for designing conservation programs (Hamrick and Godt 1996, Frankham et al. 2002). Phylogeography, an approach to study the processes that shape the geographical distribution patterns of genealogical lineages (Avise and Walker 1998, Avise 2000), can provide valuable biological information to conservation biology for endangered species, such information about genetic diversity, population structure, and Evolutionarily Significant Units (ESU), which are essential for developing useful conservation strategies and actions (Avise et al. 1987, Moritz 1994, 1995, Pope et al. 1998, Osborne et al. 2000). The use of this kind of research has become increasingly popular in biological conservation in recent years (Ge et al. 2011, Chung et al. 2013, Beatty et al. 2014). *Fritillaria pallidiflora* Schrenk, a perennial herb with extremely high pharmacological value, is naturally distributed in forests, thickets, meadows, grassy slopes, and mountain steppes at altitudes between 1100–2200 m a.s.l. in Yining, Huoqiang and Wenguan counties in the Xinjiang Province of China, and is also distributed in Kazakhstan in central Asia (Wang and Tang 1980). It is well adapted to a wet and cool climate, and resistant to coldness. Its flower is nodding and campanulate, and the tepals are pale yellow, with darker veins and some dark red spots. The bulb is ovoid or oblong-ovoid, 1–4 cm in diameter. The bulb can be used to clear fever, relieve cough, remove phlegm and decrease asthma. The chromosome number is 2n = 24 (Chen et al. 2000). Because of harvesting, the number of natural populations of *F. pallidiflora* has declined rapidly (Yin et al. 2006), and it is presently composed of many small and isolated patches (pers. obs.). It has been listed as vulnerable in the list of rare endangered endemic higher plants in Xinjiang (Yin et al. 2006). Previous studies have focused on the chemical constituents of *F. pallidiflora* (Shen et al. 2012a, 2012b, Tan et al. 2013), and no population genetic research on *F. pallidiflora* based on molecular markers has been carried out. For this study, we sampled populations across the extent of the Tianshan Mountains covering almost the entire distribution area in China, to study phylogeographic patterns and historic evolution processes.

The maternally inherited chloroplast DNA (cpDNA) in plants is thought to evolve slowly, with low recombination and mutation rates (Li and Fu 1997, Comes and Kadereit 1998), and thus it can often better trace evolutionary history and display distinct elements in the geographic distribution of a species (Avise 2000). A number of non-coding regions of cpDNA have been successfully used in phylogeographic studies (Sosa et al. 2009, Wang et al. 2009, Wu et al. 2010, Jia et al. 2011). After sifting several gene segments, we chose three cpDNA non-coding spacers, *psbA–trnH, rps16*, and *trnS–trnG*, to conduct our study.

Our aims were to, based on genetic evidence, 1) estimate levels of genetic variation within and among
populations, 2) elucidate patterns of distribution of genetic variation, and 3) identify populations with high genetic diversity and potential conservation units.

Material and methods

Plant material

A total of 150 individuals from 11 populations of *Fritillaria pallidiflora* were sampled, covering almost the entire geographic distribution in China. Eight populations were sampled in the Yili Valley, and three were sampled in Wenquan County; ten to fifteen individuals were sampled per population. To avoid clonally propagated ramets, we sampled only one individual in a cluster, and the minimum distance between clusters was 30 m. Fresh leaves were gathered from each individual and dried in silica gel.

*Fritillaria* in China has been divided into two sections, sect. *Fritillaria* and sect. *Theria*. We selected *F. walujewii*, *F. hupehensis*, *F. taipaiensis* and *F. cirrhosa* within sect. *Fritillaria* as outgroups for the phylogeny and network analysis.

DNA extraction, amplification, and sequencing

Using a modified 2×CTAB method, total genomic DNA was extracted from silica-gel-dried leaf tissue (Rogers and Bendich 1985; Doyle and Doyle 1987). The intergenic chloroplast spacer *trnH–psbA* was amplified and sequenced using the primers and protocols of Sang et al. (1997), the *trnS–trnG* spacer according to Shaw et al. (2005), and the *rps16* region according to Oxelman et al. (1997). Amplification products were first purified by PCR Product Purification Kits, and then with the forward and reverse primers of the amplification reactions, sequencing reactions were conducted with the DYEnamic ET Terminator Kit, using an ABIPRISM3730 automatic DNA sequencer. Electropherograms were edited and assembled by SEQUENCHER 4.8, aligned by the program CLUSTAL W (Thompson et al. 1994), and refined by visual inspection. Indels were coded as single binary characters (Simmons and Ochoterena 2000), and the alignments were adjusted manually. Haplotypes were identified by the program TCS (Clement et al. 2000).

Data analysis

Within-population diversity (h), total gene diversity (h), and genetic differentiation (G), as well as population subdivision for phylogenetically ordered alleles (G) were estimated using HAPLONST (http://www.pierroton.inra.fr/genetics/lab/Software/index.html). To test the spatial genetic structure of the species, spatial analysis of molecular variance was performed using the program SAMOVA (Dupanloup et al. 2002). It defines groups that are geographically homogeneous and genetically differentiated from each other; the analysis was run from K = 2 to K = 10. At completion, the number of groups maximizing the proportion of total genetic variance (FCT) was retained as the final grouping pattern. Standard diversity indices, including haplotype diversity (h; Nei 1987), mean number of pairwise differences (IT; Tajima 1983), and nucleotide diversity (IT; mean number of pairwise differences per site; Nei 1987) were calculated for each population and group using the program ARLEQUIN (Excoffier et al. 2005), and analysis of molecular variance (AMOVA) was conducted to study the genetic structure of the populations (Excoffier et al. 1992). Populations with higher levels of genetic variation were examined as genetic diversification centers, or as possible sites of refuge (Taberlet and Cheddadi 2002). Using statistical parsimony, a network of all haplotypes was constructed (Templeton et al. 1992) with a maximum connection limit equal to 40 steps, in the program TCS (Clement et al. 2000).

Phylogenetic analysis of the cp haplotypes was carried out by maximum parsimony (MP) with the program PAUP* (Swofford 2002). MP trees were constructed with a heuristic search, and 100 random additions of sequences, swapping tree–bisection–reconnection (TBR) branches, with the MULTEES, COLLAPSE, and STEEPEST DESCENT options in effect. Characters were weighted equally and their state changes were treated as unordered; gaps in sequences were treated as a fifth character state. The reliability of the MP trees was tested by 1000 bootstrap replicates. Bayesian analysis was also used to search for tree topologies and estimate divergence times among the lineages, using the program BEAST (Drummond et al. 2002, Drummond and Rambaut 2007). We used the HKY substitution model and a constant-size coalescent tree prior. Each indel was treated as a single mutation event and coded as a substitution in the Bayesian analysis (Simmons and Ochoterena 2000). For most angiosperm species, the cpDNA substitution rates have been estimated to be in the range 1.0 × 10^{-9} s^{-1} y^{-1} to 3.0 × 10^{-9} s^{-1} y^{-1} (Wolfe et al. 1987). Because of fewer variable sites, we used the lower rate 1.0 × 10^{-9} to estimate divergence time of major clades. After a burn-in of 1 000 000 steps, all parameters were sampled once every 1000 steps from 10 000 000 Markov chain Monte Carlo steps. By visual inspection of plotted posterior estimates, we examined convergence of the stationary distribution using TRACER (Drummond and Rambaut 2007), and the effective sample size for each parameter sampled was found to be over 200.

Results

Sequence analysis

The aligned sequence length for the *trnH–psbA* spacer was 286 base pairs (bp) for the *rps16* spacer 829 bp, and 678 bp for the *trnS–trnG* spacer. A total of 3 informative characters were found in the aligned sequence data: 2 nucleotide substitutions (positions 425, 949) and 1 indel (position 1695). Within the 150 sampled individuals from 11 populations, a total of 6 haplotypes (A–F) were identified (Table 1). Bank accession numbers of the cpDNA sequences are KJ956421-KJ956423, KP168448-KP168450. GenBank no. of the outgroup sequences are KJ956409, KJ956410, KJ956413, KF712486, KC543997, KF769143.

Haplotype geographical distribution

The geographic distribution of cp haplotypes and frequency of haplotypes of each population are presented in Fig. 1 and Table 1. In general, haplotype distribution is fragmented.
Table 1. Details of sample locations, sample size, and haplotype frequencies for 11 populations of *Fritillaria pallidiflora*. Numbers in parenthesis represent the number of individuals with each haplotype.

<table>
<thead>
<tr>
<th>Number</th>
<th>Regions</th>
<th>Location</th>
<th>Latitude (N)</th>
<th>Longitude (E)</th>
<th>Altitude (m)</th>
<th>Haplotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Daxigou</td>
<td>Daxigou1</td>
<td>44.42</td>
<td>80.76</td>
<td>1303</td>
<td>A(15)</td>
</tr>
<tr>
<td>2</td>
<td>Daxigou</td>
<td>Daxigou2</td>
<td>44.44</td>
<td>80.78</td>
<td>1184</td>
<td>A(15)</td>
</tr>
<tr>
<td>3</td>
<td>Daxigou</td>
<td>Daxigou3</td>
<td>44.50</td>
<td>80.81</td>
<td>1654</td>
<td>B(13), C(2)</td>
</tr>
<tr>
<td>4</td>
<td>Lucaogou</td>
<td>Lucaogou</td>
<td>44.36</td>
<td>80.99</td>
<td>1341</td>
<td>A(8), B(1), D(1), E(5)</td>
</tr>
<tr>
<td>5</td>
<td>Guozigou</td>
<td>Guozigou1</td>
<td>44.46</td>
<td>81.07</td>
<td>1909</td>
<td>E(15)</td>
</tr>
<tr>
<td>6</td>
<td>Guozigou</td>
<td>Guozigou2</td>
<td>44.48</td>
<td>81.10</td>
<td>1894</td>
<td>E(15)</td>
</tr>
<tr>
<td>7</td>
<td>Yining</td>
<td>Yining1</td>
<td>44.20</td>
<td>81.70</td>
<td>1652</td>
<td>F(15)</td>
</tr>
<tr>
<td>8</td>
<td>Yining</td>
<td>Yining2</td>
<td>44.15</td>
<td>81.73</td>
<td>1831</td>
<td>F(15)</td>
</tr>
<tr>
<td>9</td>
<td>Wenquan</td>
<td>Haxia</td>
<td>44.79</td>
<td>81.24</td>
<td>2130</td>
<td>E(10)</td>
</tr>
<tr>
<td>10</td>
<td>Tuosigou</td>
<td>Tuosigou1</td>
<td>44.86</td>
<td>81.05</td>
<td>2128</td>
<td>E(10)</td>
</tr>
<tr>
<td>11</td>
<td>Tuosigou</td>
<td>Tuosigou2</td>
<td>44.88</td>
<td>81.04</td>
<td>1730</td>
<td>B(6), E(4)</td>
</tr>
</tbody>
</table>

The Daxigou region, Guozigou region, and Yining Region did not share any haplotypes. The Lucaogou region shared haplotypes A and B with the Daxigou region, and haplotype E with the Guozigou region. The Wenquan region shared haplotype B with the Daxigou region, and haplotype E with the Lucaogou and Guozigou regions.

**Haplotype relationships**

In the network, haplotype E is ancestral (Fig. 2), and haplotypes D and F have a close relationship with E, differing only by one substitution. Haplotype A has a close relationship with D, diverging via one mutation; haplotype B is derived from A by one mutation, and haplotype C is derived from B via one mutation.

**Genetic diversity and genetic structure**

Spatial genetic analysis of cpDNA haplotypes indicated that $F_{CT}$ increased to a maximal value of 0.8499 at $K = 6$ while $K$ (the number of groups) was being raised from $K = 2$ to $K = 10$. The grouping pattern of populations corresponding to $K = 6$ was: 1) populations 1–2; 2) population 3; 3) population 4; 4) populations 5–6, 9–10; 5) populations 7–8; 6) population 11. Within-population gene diversity ($h_q$) was 0.129 (SE 0.0702), and total gene diversity ($h_T$) was 0.768 (SE 0.0710). Differentiation among populations was very high ($G_{ST} = 0.832, SE 0.0935$), indicating a significant population structure in *F. pallidiflora*. As shown by AMOVA analysis, 81.69% ($p < 0.001$) of the total variation occurred among populations, and only 18.31% occurred within populations. Based on the SAMOVA groups, AMOVA showed that 85.00% ($p < 0.001$) of the total variation occurred among groups (Table 2). Among all the populations, population 4 (group 3) had the highest haplotype diversity, mean number of pairwise differences, and nucleotide diversity.

**Phylogenetic analysis and divergence time between main clades**

In MP analysis, one tree was retained. The tree has the same topology as that produced by Bayesian analysis, and so we present only the Bayesian tree (Fig. 3). In this tree, *F. pallidiflora* is resolved as monophyletic, and the relationship is well supported (100% bootstrap support and 1.00

![Figure 1. Geographical distribution of *F. pallidiflora* in China. Population numbers correspond to those in Table 1. The rough distribution range of *F. pallidiflora* in China is indicated by the ellipse.](image-url)
posterior probability). Most of the remainder of the clades are also well supported. Haplotypes A, B, C and D, distributed in the Daxigou and Lucaogou regions, cluster into a clade with a moderately high support value (63%; 1.00), and in this clade, haplotypes B and C cluster into a subclade with a high support value (63%; 1.00). Divergence times between the major lineages of *F. pallidiflora* (nodes 1, 2 in Fig. 3) are estimated as 0.82 Ma ((0.26, 1.59), the Maximum Glaciation in the early Pleistocene), and 0.37 Ma ((0.06, 0.84), the Penultimate Glaciation in the middle Pleistocene), respectively, according to Shi et al. (2005).

**Discussion**

**Genetic variation of three non-coding spacers of cpDNA in *F. pallidiflora***

Total gene diversity ($h_T = 0.768$) in *Fritillaria pallidiflora* is found to be high. The species has probably survived in northwestern China since the Tertiary (Wu et al. 2010), and its long presence in the area has evidently promoted the accumulation of a significant degree of genetic diversity (Yuan et al. 2008, Falchi et al. 2009). In addition, the varied habitats occupied by *F. pallidiflora* may harbor locally adapted gene variants as a consequence of differential geology and topography across the species distribution range.

Within-population gene diversity ($h_S = 0.129$) is relatively low compared with total gene diversity ($h_T$), resulting in a high level of differentiation among populations ($G_{ST} = 0.832$). The high level of differentiation among populations and SMOVA groups is also supported by AMOVA analyses (Table 2). Based on our observations, individuals of *F. pallidiflora* usually cluster in the wild, and thus we speculate that because of gravity, the seeds aggregate and germinate around the parent plant (Van der Pijl 1969), which promotes inbreeding among individuals within populations. Geographic barriers may explain the genetic differentiation among populations. The Tianshan Range in China contains more than twenty east-west mountains and valleys, and the altitude of the main mountains exceeds 4000 m a.s.l. (Wei and Hu 1990). The distribution of *F. pallidiflora* is therefore fragmented by many valleys of different shapes and sizes. These multiple deep valleys may obstruct gene flow and increase genetic differentiation among populations, promoting the inbreeding probably responsible for the observed high homozygosity within populations (Hamrick and Godt 1989).

**Allopatric divergence in *F. pallidiflora***

Climate oscillation during the Quaternary is usually considered an important factor influencing current geographical distribution patterns and population genetic structures of species (Hewitt 2004). The Tianshan Mountains have been repeatedly glaciated in the past, and the extent of glacial area has varied in response to the alternation of ice ages and interglacials (Shi et al. 2005). In 0.8–0.6 Ma, glaciers in high mountains of northwest China reached the maximum (Shi et al. 2005). Extremely low temperatures and aridity during glacial periods might thus be correlated with the intraspecific differentiation in allopatric regions for the species. The lineages of *F. pallidiflora* (node 1 in Fig. 3) diverged in the Maximum Glaciation in the early Pleistocene; and those of node 2 diverged during the Penultimate Glaciation. We speculate that an extremely harsh climate decreased the viability of the species, reduced and fragmented its distribution range, and thus promoted the accumulation of the genetic differentiation among isolated populations or regions. Similar phenomena appear to have occurred also in other species, such as *Hippophae tibetana* and *Aconitum gymnandrum* (Wang et al. 2009, Jia et al. 2011).

**Conservation implications for *F. pallidiflora***

Genetic drift easily occurs in endangered species with small population sizes and narrow distributions, especially when
gene flow between populations is restricted (Frankham et al. 2002). Consequently, this may reduce the genetic diversity of a species. Another result of small fragmented populations is sensitivity to inbreeding, which reduces heterozygosity and fitness-related genetic variation, and thus increases the risk of extinction (Fischer and Matthies 1997, Reed and Frankham 2003). Given the small sizes and fragmented distributions of natural populations of *F. pallidiflora*, the maintenance of genetic diversity is critical for its long-term survival (Frankel 1983). In *F. pallidiflora*, for the purpose of retaining the existing genetic diversity, reservation populations covering major genetic variations should be established. As shown by the haplotype network (Fig. 2), haplotype E is ancestral, and for comparison of population genetic diversity, population 4 in Lucaogou includes the main genetic diversity of *F. pallidiflora*, (Table 1, 3), and thus we suggest the Lucaogou region as the possible site of a glacial refugium or a diversification center (Taberlet and Cheddadi 2002), and it should be the area of greatest emphasis in conservation.

Habitat destruction and fragmentation would inevitably increase the risk of extinction (Ledig et al. 2002, Nakagawa 2004). Today, many natural habitats of *F. pallidiflora* in the Tianshan Mountains have been destroyed due to human overexploitation. In order to develop effective strategies for conservation, Evolutionarily Significant Units (ESU) need to be defined. Recognizing ESUs as reciprocally monophyletic lineages within species ensures that different lineages can be managed separately and thus high genetic diversity can be maintained (Crandall et al. 2000). In *F. pallidiflora*, we identified ESUs by reciprocal monophyly (Frankham et al.)

Figure 3. Phylogenetic relationships of six haplotypes of *F. pallidiflora* and related species. Numbers above branches are support values (maximum parsimony bootstrap values greater than 50%), and numbers below branches are posterior probabilities greater than 0.8). The estimated divergence time at node 1 is about 0.82 Mya, at node 2 is about 0.37 Mya.

### Table 3. Measures of haplotype diversity, mean number of pairwise differences, and nucleotide diversity within sampled locations and SAMOVA groups of *Fritillaria pallidiflora* based on three sequences.

<table>
<thead>
<tr>
<th>Population</th>
<th>n</th>
<th>Haplotype diversity</th>
<th>Mean number of pairwise differences</th>
<th>Nucleotide diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>0.2476</td>
<td>0.2476</td>
<td>0.0825</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>0.6381</td>
<td>1.1238</td>
<td>0.3746</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>7</td>
<td>15</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>11</td>
<td>10</td>
<td>0.5333</td>
<td>1.6000</td>
<td>0.5333</td>
</tr>
</tbody>
</table>

SAMOVA groups

<table>
<thead>
<tr>
<th>Groups</th>
<th>n</th>
<th>Haplotype diversity</th>
<th>Mean number of pairwise differences</th>
<th>Nucleotide diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Group 2</td>
<td></td>
<td>0.2476</td>
<td>0.2476</td>
<td>0.0825</td>
</tr>
<tr>
<td>Group 3</td>
<td></td>
<td>0.6381</td>
<td>1.1238</td>
<td>0.3746</td>
</tr>
<tr>
<td>Group 4</td>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Group 5</td>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Group 6</td>
<td></td>
<td>0.5333</td>
<td>1.6000</td>
<td>0.5333</td>
</tr>
</tbody>
</table>
A monophyletic gene genealogy of cp haplotypes was recovered in the MP tree (Fig. 3), thus we identified the regions Daxigou and Lucaogou as one ESU, and regions Guozigou, Wenquan, and Yining as the other ESU. In conclusion, there is a significant genetic differentiation among populations in *F. pallidiflora* based on the cpDNA spacers we selected, and we speculate that low temperatures and aridity during glacial periods caused range shrinkage and fragmentation, and intraspecific differentiation in allopatric regions. For setting a conservation management plan, we identified one population with the highest genetic diversity, in the Lucaogou region, and two ESUs in the whole distribution area in China. To obtain a better understanding of the genetic structure and evolutionary history of the species, and provide additional useful information for conservation, further studies by other molecular techniques will be needed in the future.

Acknowledgements – This research was supported by grants from the Western Doctor Project (XBSBS201306) of Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences. We thank Huiliang Liu, Xiaobin Zhou, Xiaoshan Kang at the Xinjiang Institute of Ecology and Geography, CAS, for their kindly help in sampling in the wild.

References


Yin, L.-K. et al. 2006. Rare endangered endemic higher plants in Xinjiang of China. – Science Press.