The LINNEAN SOCIETY of London BIOLOGICAL Journal of the Linnean Society

Biological Journal of the Linnean Society, 2013, ••, ••–••. With 3 figures

Local genetic structure in the critically endangered, cave-associated perennial herb *Primulina tabacum* (Gesneriaceae)

ZHENG-FENG WANG†, HAI REN†, ZHONG-CHAO LI, QIAN-MEI ZHANG, KAI-MING LIANG, WAN-HUI YE* and ZHANG-MING WANG

Key Laboratory of Vegetation Restoration and Management of Degraded Ecosystems, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China

Received 11 January 2013; revised 19 February 2013; accepted for publication 20 February 2013

The local spatial genetic structures of cave-associated plants are seldom studied. Given that these plants are mainly confined to small areas in and around the entrances of caves, we hypothesized that they might lack genetic structures at local scales. To test this hypothesis, we sampled two large populations (named D and T) of a critically endangered perennial herb, *Primulina tabacum*, which is endemic to karst caves in southern China. We analysed nine microsatellite loci and sequenced four chloroplast DNA (cpDNA) intergenic spacer regions to study the genetic diversity and structure within and between both populations. Both populations have distinct genetic characteristics. Samples from two subpopulations in population D showed considerable genetic divergence. This is not consistent with the hypothesis that *P. tabacum* has a weak genetic structure at a local scale. However, 94% of the individuals in population T shared the same multilocus genotype, which indicates little genetic structure within this population. The contributions of seed flow, pollen flow and (sub)population history to the genetic diversity and structure in each and both populations are discussed. Our study is the first to investigate local genetic diversity and structure in a cave-associated plant, and provides valuable information for the sustainable conservation of such species. *Biological Journal of the Linnean Society*, 2013, ••, ••–••.

ADDITIONAL KEYWORDS: asymmetric gene flow – breeding system – cave plant – conservation genetics – cpDNA – genetic diversity – limestone – microsatellite – selfing – wind-borne seed.

INTRODUCTION

A cave is a unique environment for species to inhabit (Pipan & Culver, 2012). In general, caves are geographically discontinuous habitats. Consequently, owing to their specific environmental requirements, many cave-associated species are fragmented, have a restricted distribution and are endemic to the caves they inhabit (Christman *et al*., 2005; Schilthuizen, Rutten & Haase, 2012). Unlike obligate cave-dwelling species, cave-associated plants mainly grow in and around the entrances of caves in full or diminished sunlight (Glime, 2007) and are not found deep within the cave in complete darkness. Consequently, cave-associated plants are expected to be able to exchange genes frequently with individuals that grow outside the cave. Accordingly, cave-associated plants display limited local genetic differentiation. Although certain studies of cave-dwelling animals have indicated that highly mobile species, such as the bat *Miniopterus schreibersii* (Pereira *et al*., 2009) and the fish *Astyanax mexicanus* (Avise & Selander, 1972), show such patterns, other studies have reported medium to high levels of genetic isolation for some highly mobile cave-dwelling animals because of the complex cave systems that are inhabited by the species (Renno *et al*., 2007; Flot, Wörheide & Dattagupta, 2010; Bradic *et al*., 2012).

*Corresponding author. E-mail: why@scbg.ac.cn
†Joint first authors.

Thus, the spatial genetic structure of a cave-dwelling species might differ among species and between caves.

Primulina tabacum Hance is a calciphilous perennial herb that belongs to the Gesneriaceae. In general, its distribution is restricted to the entrances of karst cave drainages along the border of northern Guangdong, China. Its population sizes have decreased drastically during the last three decades mainly because of increasing anthropogenic disturbances, such as increased exploitation by tourism operators and the excavation of limestone. In 1999, *P. tabacum* was placed on the list of 'First Class Protected Key Wild Plants of China' (Peng & Cheng, 2002).

Primulina tabacum grows mainly in cave entrances, with a continuous distribution from the exterior to the interior of a cave (Supporting Information Fig. S1). In addition, its seeds are tiny and readily dispersed by air currents. Thus, the seeds of *P. tabacum* may be spread throughout the areas outside and inside a cave entrance, including the cave walls and rock clefts. All of these factors suggest that local genetic structure is unlikely in *P. tabacum*. Ni *et al*. (2006) sampled 24 individuals from each of four *P. tabacum* populations located tens of kilometres apart, and used dominant amplified fragment length polymorphism (AFLP) molecular markers to study the genetic diversity in the populations. However, owing to the limited sample sizes in each population, their study was not able to provide information on genetic structure at a local scale within populations.

Given that a knowledge of the local spatial genetic structure of a species is important when designing effective conservation strategies (Escudero, Iriondo & Torres, 2003), the aim of the present study was to examine such structure in *P. tabacum* using both microsatellite and chloroplast DNA (cpDNA) markers to: (1) investigate genetic diversity and differentiation at a fine spatial scale and to test the hypothesis that there is no local genetic structure; and (2) to suggest specific conservation strategies for *P. tabacum* on the basis of the results. We are unaware of any previous studies that have focused on the patterns of local spatial genetic structure for *P. tabacum* specifically, and for cave plants in general.

MATERIAL AND METHODS

Study species

Primulina tabacum is a stemless perennial herb. Its leaves are oval or cordiform, and they are borne on rhizomes, form a rosette and have the odour of tobacco. The flowers are violet and tubular in shape, with a corolla that has a diameter of approximately 1.6 cm. The species flowers from August to October. *Primulina tabacum* is a hermaphroditic and animal-pollinated species (S.-J. Li, South China Botanical Garden, Chinese Academy of Sciences, pers. comm.). The corolla tube is narrow (9 mm long and 3 mm in diameter at the mouth), which prevents large insects from penetrating the tube (Ni *et al*., 2006). *Primulina tabacum* is self-compatible (Li *et al*., 2006) and does not propagate clonally in the field (S.-J. Li, pers. comm.). Its seeds are ellipsoidal, approximately 0.4 mm in length, and are dispersed easily by wind, because plants often grow on cave walls and in external rock clefts near cave entrances. *Primulina tabacum* prefers low light, relatively wet soil and humid habitats (Ren *et al*., 2010a, b). In the field, moss plays a key nurse role for the establishment of *P. tabacum* plants (Ren *et al*., 2010a).

Primulina tabacum has a very restricted distribution and has only been found in the limestone area of northern Guangdong and southern Hunan provinces, China, at approximately 300 m above sea level (Ren *et al*., 2010b). Specimens of *P. tabacum* were first collected by Andrew Henry in 1881 from Lianzhou, Guangdong province (Hance, 1883). It was considered to be extinct after 1936 (Ni *et al*., 2006) until its existence was reported again in 1986 and 1991 (Ni *et al*., 2006; Ren *et al*., 2010b). Based on our field investigation, *P. tabacum* now only exists in five locations in China, two of which are in Hunan province and three of which are in Guangdong province (refer to detailed location information in Ren *et al*., 2010b and fig. 1 therein). However, recent taxonomic analysis has indicated that the individuals in Hunan province should belong to a new species (S.-J. Li, pers. comm.), which is a sister of *P. tabacum*. Therefore, although we collected leaf samples from the two locations in Hunan province, we did not use them in this study.

Sample sites and leaf collection

Primulina tabacum leaf samples were collected from the three remaining locations in Guangdong province, with one in Qingyuan City and the other two in Liangzhou City. However, samples from Qingyuan City were lost during their transfer from an old laboratory to a new one. Given that the population in Qingyuan City was small (with only ten individuals that could be collected from it), we did not resample these plants to avoid disturbing the population. The two populations in Liangzhou City are located at the caves Dixiahe (112°21′E, 25°0′N) and Tiantangdong (112°30′E, 25°1′N). Accordingly, we refer to them as populations D and T, respectively. The two populations are found about 15 km apart and have similar

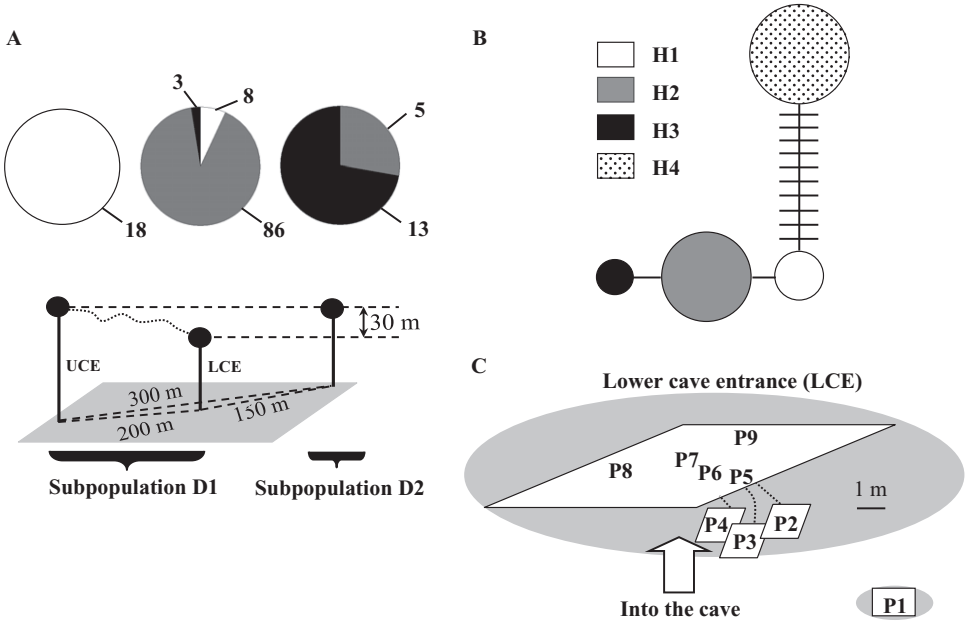

Figure 1. A, The orientations of the caves in which *Primulina tabacum* population D is located. UCE, upper cave entrance; LCE, lower cave entrance. The dotted line represents a tunnel connecting UCE and LCE. The pie diagrams summarize the number of individuals with different haplotypes in the UCE, LCE and subpopulation D2. The haplotypes are defined in (B). B, Statistical parsimony network of chloroplast DNA (cpDNA) haplotypes (H1–H4) in *P. tabacum*. Horizontal lines represent mutational steps that separate haplotypes, and a gap is treated as one mutational step. The sizes of the circles are proportional to the number of individuals that belong to the respective haplotypes. C, Patches in which *P. tabacum* individuals in the lower cave entrance were sampled. The dotted lines represent tunnels connecting the patches.

environments in general (Ren *et al*., 2010b). Ren *et al*. (2010b) used a nearby village name of Shangbochang to name population T.

Population D was divided into two subpopulations: D1 and D2. Subpopulation D1 was located in a cave with two open ends that are connected by a subterranean river of 1500 m in length. Both ends of the cave face north, and each has a single entrance; the upstream and downstream entrances are referred to as the upper and lower cave entrances, respectively (Fig. 1). The hall of the upper entrance is approximately 100 m in length, 100 m in width and 85 m in height; the hall of the lower entrance is approximately 60 m in length, 30 m in width and 40 m in height. This cave has been a popular tourist attraction since 1988. Subpopulation D2 was located in a small (length, 60 m; width, 30 m; height, 30 m) cave beside D1.

Population T was located in a cave that has not been exploited by tourism. The length of the cave is unknown. This cave is visited only occasionally by local villagers at the entrance, and the entrance hall is approximately 20 m in length, 15 m in width and 15 m in height.

Given that *P. tabacum* individuals were distributed patchily within the two populations, leaf samples were collected and separated on the basis of the patches within a population or subpopulation (Fig. S1). Nine patches (P1–P9) were located in subpopulation D1, and six patches (P10–P15) in population T. These patches, which were numbered from the smallest to the largest in D1 and T, corresponded to patch positions from outside the cave entrance to inside the caves. All leaf samples were collected in the field and dried in silica gel in sealed plastic bags prior to DNA extraction.

Genetic analysis

Total genomic DNAs were extracted from the dried leaves of *P. tabacum* using a modified cetyltrimethylammonium bromide (CTAB) method (Doyle, 1991). Nine nuclear microsatellite loci (Table 1) were analysed using the amplification procedures of Wang *et al*. (2009). Four cpDNA intergenic spacer regions (*atpB–rbcL*, *trnL*^{UAA}*–trnF*^{GAA}, *trnD*^{GUC}*–trnT*^{GGU} and *trnS*^{GCU}*–trnG*^{UCC}) were used. The primers used for PCR amplification and sequencing of *atpB-rbcL* (F, 5′-TCCTAATAATTGTTGTACCTCACA-3′; R, 5′-AACACCAGCTTTAAATCCAACA-3′) were published by Ebert & Peakall (2009), those for *trnL*^{UAA}*–trnF*^{GAA} (F, 5′-CGAAATCGGTAGACGCTACG-3′; R, 5′-ATTTGAACTGGTGACACGAG-3′) by Taberlet *et al*.

Table 1. Genetic variation for nine microsatellite loci in two *Primulina tabacum* populations, including the total number of alleles (N_A), expected heterozygosity (H_E), observed heterozygosity (H_O) and the inbreeding coefficient (F_{IS}). Results without multilocus repeats are presented in parentheses. All F_{IS} values are significant at the 0.01 level

	Population D				Population T			
Locus	N_A	H_E	H_O	F_{IS}	N_A	H_E	H_O	F_{IS}
Pta01	8	0.757 (0.704)	0.167 (0.220)	0.781 (0.689)	2	0.089 (0.533)	0.000 (0.000)	1.000 (1.000)
Pta02	4	0.338 (0.397)	0.180 (0.218)	0.467 (0.452)	1	0.000 (0.000)	0.000 (0.000)	—— (——)
Pta03	3	0.623 (0.551)	0.135 (0.178)	0.784 (0.677)	1	0.000 (0.000)	0.000 (0.000)	—— (——)
Pta04	7	0.802 (0.800)	0.250 (0.310)	0.689 (0.614)	1	0.000 (0.000)	0.000 (0.000)	—— (——)
Pta05	7	0.609 (0.600)	0.195 (0.238)	0.680 (0.605)	2	0.018 (0.533)	0.000 (0.000)	1.000 (1.000)
Pta06	3	0.418 (0.396)	0.158 (0.208)	0.623 (0.477)	1	0.000 (0.000)	0.000 (0.000)	—— (——)
Pta07	3	0.213 (0.257)	0.038 (0.050)	0.824 (0.808)	1	0.000 (0.000)	0.000 (0.000)	—— (——)
Pta08	17	0.834 (0.842)	0.250 (0.307)	0.701 (0.633)	1	0.000 (0.000)	0.000 (0.000)	—— (——)
Pta10	8	0.687 (0.741)	0.298 (0.364)	0.568 (0.511)	1	0.000 (0.000)	0.000 (0.000)	—— (——)

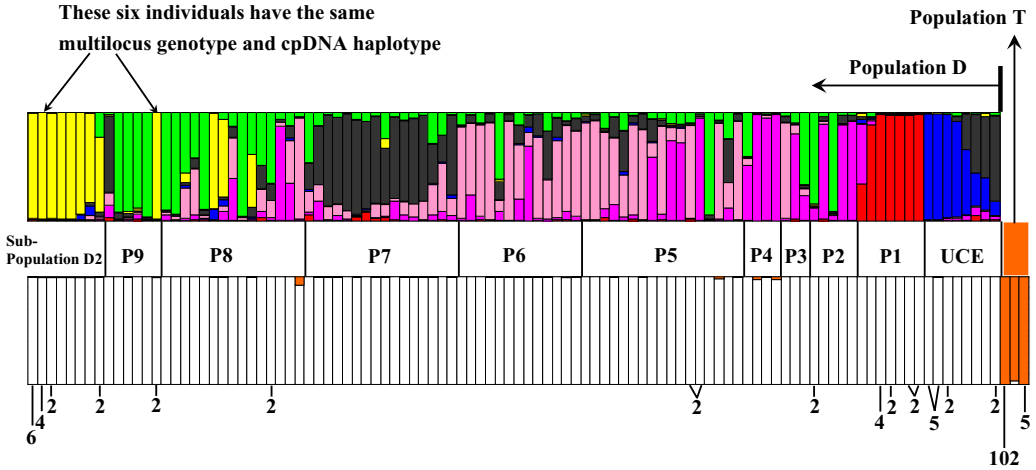

Figure 2. Coloured bars representing the assignment probability that each multilocus genotype belongs to a particular cluster in the STRUCTURE analysis. Upper coloured bars represent seven clusters for individuals from population D; lower coloured bars represent two clusters for individuals from both populations D and T. The numbers below the bars indicate the numbers of individuals with the same multilocus genotype, and only genotypes with more than one individual are indicated. P numbers denote patches (Table 2). UCE, upper cave entrance.

(1991), those for $trnD^{GUC}$–$trnT^{GGU}$ (F, 5′-ACCAATTGA ACTACAATCCC-3′; R, 5′-CTACCACTGAGTTAAAAG GG-3′) by Demesure, Sodzi & Petit (1995), and those for $trnS^{GCU}$–$trnG^{UCC}$ (F, 5′-GCCGCTTTAGTCCACTCA GC-3′; R, 5′-GAACGAATCACACTTTTACCAC-3′) by Shaw *et al*. (2005). The PCR amplifications of the four cpDNA regions were performed in 20-μL reaction mixtures, and consisted of 5 ng of template DNA, 50 mM KCl, 20 mM Tris-HCl (pH 8.0), 1.5 mM $MgCl_2$, 0.5 μM of each primer, 0.6 mM of each deoxynucleoside triphosphate (dNTP) and 0.5 U of *Taq* DNA polymerase (Takara Bio Inc.). The reaction protocol was as follows: 4 min at 94 °C, followed by 35 cycles of 94 °C for 30 s, 60 °C for 60 s and 72 °C for 2 min, and then, finally, 20 min at 72 °C. The PCR products were sequenced using an ABI 3730XL automated sequencer.

Sequences of the four cpDNA regions were obtained for all 133 individuals in population D. Given that only three multilocus genotypes were identified among the 108 plants of population T – which comprised one, five and 102 plants, respectively (Fig. 2) – we analysed the six individuals with the two least frequent genotypes and five of the 102 individuals with the predominant genotype to examine possible variation in their cpDNA haplotypes. The results showed that all 11 individuals shared the same cpDNA haplotype. Consequently, we did not sequence the remainder of the 108 individuals in population T, but assumed that most of the individuals probably

arose by selfing, and thus shared the same cpDNA haplotype (see Results and Discussion sections).

Data analyses

The genetic variation parameters of observed heterozygosity (H_O), unbiased expected heterozygosity (H_E) (Nei, 1978) and the inbreeding coefficient (F_{IS}) were estimated using GENETIX 4.05 (Belkhir *et al.*, 1996–2004). Deviations in genotypic patterns from Hardy–Weinberg equilibrium and linkage disequilibrium were tested using GENEPOP 4.0.7 (Rousset, 2008).

Population genetic structure was analysed using Bayesian individual-based clustering methods implemented in STRUCTURE 2.3.1 (Pritchard, Stephens & Donnelly, 2000; Falush, Stephens & Pritchard, 2003). Using a model that assumed admixture and correlated allele frequencies, 20 independent runs were performed for each K (putative cluster numbers from 1 to 5), with 10^6 iterations after a burn-in period of 10^6 generations. The optimal K was then determined by the analysis of ΔK values (Evanno, Regnaut & Goudet, 2005). CLUMPP 1.1.2 (Jakobsson & Rosenberg, 2007) was used to align the results of the 20 runs for the inferred K. Given that the two clusters that were revealed by STRUCTURE (see Results) only differentiated populations D and T, we excluded population T and repeated the analysis with K values that ranged from 2 to 11 for population D. The optimal K was also determined by the analysis of ΔK values. Relationships among individuals of population D were also visualized using principal component analysis (PCA), as implemented using GENALEX 6.3 (Peakall & Smouse, 2006).

The maximum likelihood method implemented in MIGRATE-N 3.1.6 (Beerli & Felsenstein, 1999, 2001) was used to estimate historical gene flow (N_em, effective number of migrants per generation, where N_e is the effective number of individuals and m is the migration rate) between the upper cave entrance, lower cave entrance and subpopulation D2 in population D on the basis of microsatellite genotyping data. Based on an infinite allele model, the estimated N_em was the mean of five independent replicates, with each including 30 short chains (2×10^5 trees sampled), four long chains (10^6 trees sampled), each with increments of 100 steps, a static heating scheme with four chains (1.0, 1.5, 3.0 and 10 000.0) and a swapping interval of one.

Recent gene flow between the upper cave entrance, lower cave entrance and subpopulation D2 within population D was estimated using the individual assignment-based method BAYESASS 1.3 (Wilson & Rannala, 2003). The estimated result was the mean of three independent runs with each including 10^7 iterations and the first 10^6 iterations discarded as burn-in.

Sequences were checked manually and aligned using BIOEDIT 7.0.9.0 (Hall, 1999). All unique sequences used in the present study were deposited in the EMBL nucleotide sequence database, with accession numbers HE799646 and HE799647 for the *atpB–rbcL* region; HE799648, HE799649, HE799650 and HE799651 for the $trnL^{UAA}$–$trnF^{GAA}$ region; HE799652 and HE799653 for the $trnD^{GUC}$–$trnT^{GGU}$ region; and HE799654 and HE799655 for the $trnS^{GCU}$–$trnG^{UCC}$ region. Genealogical relationships among cpDNA haplotypes were examined by the construction of a statistical parsimony network with TCS 1.21 (Clement, Posada & Crandall, 2000) with a 95% connection limit and treating gaps as a fifth state.

RESULTS

Genetic variation at nuclear microsatellite loci

One allele was detected at each of seven microsatellite loci and two alleles at each of the other two loci in population T (Table 1). The number of alleles ranged from three to 17 among the nine microsatellite loci in population D, and all loci showed a significant heterozygote deficit, as indicated by the positive F_{IS} values (Table 1). No consistently significant patterns of linkage disequilibrium were detected for any locus pairs within populations after Bonferroni correction.

Genetic variation within and among populations

Population D contained a higher level of genetic variation than population T (Table 2). Only three multilocus genotypes were detected in population T, and 102 of the 108 individuals possessed the same multilocus genotype (Fig. 2). The F_{IS} values were 0.684 and 1.000 for populations D and T, respectively, and ranged from 0.264 to 0.970 among patches in subpopulation D1 (Table 2).

When all sampled individuals were analysed together, $\ln Pr(X|K)$ values increased as K increased from 1 to 5 (with values of –5604.790, –2972.215, –2481.940, –2357.975 and –2235.550, respectively) and the maximum ΔK was at $K = 2$. Thus, two was the best estimate of the number of clusters, and populations D and T formed two unique clusters (Fig. 2). Given that these two clusters provided little information about within-population genetic structure, we analysed population D alone. This analysis indicated that the preferred number of clusters was seven. At $K = 7$, subpopulation D2, the upper cave entrance, P1 and the remainder of population D were the most separated among the seven clusters (Fig. 2), and were also well separated by PCA (Fig. 3).

Table 2. Genetic variation in *Primulina tabacum* populations D and T and their patches (P1–P15). *N*, sample size; *G*, number of observed multilocus genotypes; *G*/*N*, genotypic diversity; N_A, total number of alleles; N_P, number of private alleles; H_E, expected heterozygosity; H_O, observed heterozygosity; F_{IS}, inbreeding coefficient. Results without multilocus repeats are indicated in parentheses

Sample location	Population size	*N*	*G*	*G*/*N*	N_A	N_P	H_E	H_O	F_{IS}
Population D		**133**	**101**	0.759	**60**	**38**	**0.587 (0.588)**	**0.186 (0.233)**	**0.684** (0.605**)**
Subpopulation D1	> 400	115	94	0.817	60		0.587 (0.581)	0.209 (0.245)	0.645** (0.579**)
Lower cave entrance									
P1	13	13	7	0.538			0.171 (0.225)	0.051 (0.095)	0.709** (0.596**)
P2	6	6	5	0.833			0.367 (0.385)	0.074 (0.089)	0.813** (0.789**)
P3	3	3	3	1.000			0.496 (0.496)	0.333 (0.333)	0.379* (0.379*)
P4	4	4	4	1.000			0.306 (0.306)	0.139 (0.139)	0.583** (0.583**)
P5	≈ 60	19	17	0.895			0.553 (0.564)	0.280 (0.286)	0.501** (0.500**)
P6	≈ 40	13	13	1.000			0.591 (0.591)	0.440 (0.440)	0.264** (0.264**)
P7	≈ 60	16	16	1.000			0.510 (0.510)	0.333 (0.333)	0.354** (0.354**)
P8	≈ 60	16	15	0.938			0.505 (0.508)	0.257 (0.237)	0.499** (0.542**)
P9	7	7	6	0.857			0.444 (0.428)	0.111 (0.130)	0.765** (0.717**)
Upper cave entrance	18	18	8	0.444			0.405 (0.454)	0.012 (0.028)	0.970** (0.942**)
Subpopulation D2	18	18	8	0.444	16		0.192 (0.224)	0.037 (0.056)	0.812** (0.765**)
Population T	> 400	**108**	**3**	0.028	**11**	**4**	**0.012 (0.119)**	**0.000 (0.000)**	**1.000** (1.000*)**
P10	≈ 30	17	1	0.059			0.000 (0.000)	0.000 (0.000)	——
P11	≈ 60	24	1	0.042			0.000 (0.000)	0.000 (0.000)	——
P12	≈ 40	20	2	0.100			0.011 (0.074)	0.000 (0.000)	1.000** (1.000*)
P13	≈ 40	16	2	0.125			0.049 (0.074)	0.000 (0.000)	1.000** (1.000*)
P14	10	10	1	0.100			0.000 (0.000)	0.000 (0.000)	——
P15	≈ 60	21	1	0.048			0.000 (0.000)	0.000 (0.000)	——

*$P < 0.05$; **$P < 0.01$.

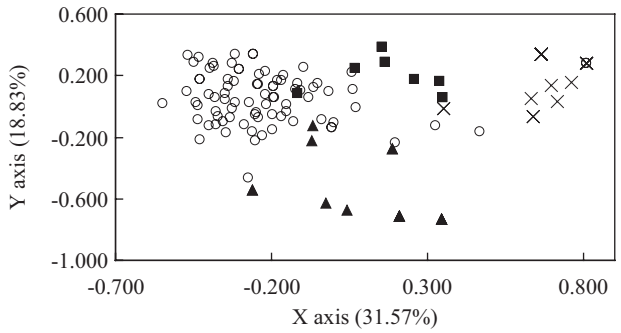

Figure 3. Principal component analysis for individuals in population D. Crosses, triangles, squares and circles represent individuals from subpopulation D2, upper cave entrance, P1 and the remainder, respectively.

The rates of historical gene flow estimated by N_em from the upper cave entrance to the lower cave entrance and to subpopulation D2 were 2.579 and 3.397, respectively. These estimates were greater than those of 0.721 and 0.0319, which were obtained for gene flow from the lower cave entrance and subpopulation D2, respectively, to the upper cave entrance (Table 3). Such asymmetrical gene flow was also observed between subpopulation D2 and the lower cave entrance: N_em was 0.979 from subpopulation D2 to the lower cave entrance, but was 0.000 in the opposite direction. Recent gene flow was low among the three locations; the highest gene flow was from subpopulation D2 to the lower cave entrance (0.069), which was more than eight-fold higher than the lowest estimate (0.008) of gene flow in the opposite direction (Table 3).

The lengths of the aligned *atpB–rbcL*, *trnL*^{UAA}–*trnF*^{GAA}, *trnD*^{GUC}–*trnT*^{GGU} and *trnS*^{GCU}–*trnG*^{UCC} sequences (including gaps) were 933, 905, 966 and 695 bp, respectively. These four regions displayed low levels of genetic variability, with one variable site in *atpB–rbcL* and four in the other three regions (Table 4). The sequence data were then combined into a dataset that comprised 3499 bp, including gaps. In total, four chloroplast haplotypes were identified among the samples, three (H1, H2 and H3) of which were present in population D, and one (H4) of which was present in population T. The mean pairwise genetic distance was very low at 0.001.

Table 3. Historical and recent (in parentheses) gene flow (N_em) among the upper and lower cave entrances of subpopulation D1 and subpopulation D2 in population D

	Recipient location		
Source location	Upper cave entrance	Lower cave entrance	Subpopulation D2
Upper cave entrance		2.579 (0.005)	3.397 (0.008)
Lower cave entrance	0.721 (0.008)		0.000 (0.008)
Subpopulation D2	0.032 (0.009)	0.979 (0.069)	

Table 4. Plastid DNA haplotypes in *Primulina tabacum* populations D (H1–H3) and T (H4). The sign '–' denotes codon deletion. Numbers below chloroplast DNA (cpDNA) region names refer to the polymorphic site position for the aligned sequences of the four regions

	cpDNA region												
	atpB-rbcL	$trnL^{UAA}$–$trnF^{GAA}$				$trnD^{GUC}$–$trnT^{GGU}$				$trnS^{GCU}$–$trnG^{UCC}$			
Haplotype	225	053	188	287	672	269	420	751	753	149	357	374	433
H1	A	I1	T	I2	C	A	–	T	G	C	–	A	C
H2	A	–	T	I2	C	A	–	T	G	C	–	A	C
H3	A	–	T	I2	T	A	–	T	G	C	–	A	C
H4	T	I1	A	–	C	G	A	G	A	T	I3	T	A

I1, CCATCCTC (8-bp insertion); I2, AATTAGAA (8-bp insertion); I3, ATATT (5-bp insertion).

The estimated statistical parsimony network is shown in Figure 1B. About ten mutational steps occurred from H4 to H1–H3, and one to two steps occurred from H1 to H3. Thus, the genetic difference was greater between H4 and H1–H3 than between any two haplotypes of H1–H3. Approximately 89% (86 of 97) of the individuals growing at the lower cave entrance of population D carried H2, whereas individuals growing at the upper cave entrance carried only H1 (Fig. 1A). Three individuals found in the lower cave entrance, one in P8, and two in P9, carried H3. The two individuals in P9 and four individuals in subpopulation D2 had identical microsatellite multilocus genotypes and H3 cpDNA haplotypes (Fig. 2).

DISCUSSION

The two populations of *P. tabacum* show distinct characteristics. One population, population T, supports the hypothesis that *P. tabacum* has a weak genetic structure at a local scale; the other, population D, does not. However, given that we detected only three multilocus genotypes and one of these was shared by more than 94% of the individuals in population T, the lack of genetic structure in this population is more likely to be caused by low genetic diversity (Table 2) than by extensive gene flow within the population, as we had hypothesized. If this had not been the case, we would find low structure in population D. In fact, except for those patches that contained only one genotype, 11 of the 16 individuals sampled in patch P13 of population T had the same genotype, and five contained the same different genotype, with no interbreeding between the two groups. This observation and our analysis of population D suggest that *P. tabacum* individuals can be effectively isolated even at a small spatial scale.

Seed of *P. tabacum* is wind dispersed, and winds can occur frequently and easily because of temperature differences between the outside and inside of the cave entrance. Considering the limited distribution areas for *P. tabacum*, seed flow within the populations cannot be restricted. Indeed, the spread of wind-borne seeds to a very large area outside and inside the cave entrance decreases the spatial genetic structure within populations. However, compared with average respective within-population F_{IS} values of 0.88, 0.15 and 0.03 (calculated with the formula $1 - H_O/H_E$, with H_O and H_E derived from Nybom, 2004) found in species with the selfing, mixed and outcrossing breeding systems, respectively, we observed a high selfing rate ($F_{IS} = 0.684$) in population D and complete selfing in T ($F_{IS} = 1.000$, Table 2). Therefore, limited pollen flow and selfing are the most likely reasons for the isolation. This proposal is consistent with the failure of one of our colleagues, S.-J. Li, to observe any

pollinator visits in one *P. tabacum* flowering season at the upper cave entrance in population D after 200 h of observation, and the resulting conclusion that *P. tabacum* could be self-pollinated by corolla abscission when pollinators are scarce (Li *et al*., 2006). Moreover, Martén-Rodríguez & Fenster (2010) found that, in 13 Gesnerieae species, self-pollination was frequent in tubular-shaped flowers that required specialized pollinators. Given that *P. tabacum* has a typical tubular-shaped flower with a narrow corolla mouth, its pollinators might also be specialized, and limited outcrossing among the individuals of *P. tabacum* could be caused by a lack of pollinators. In fact, selfing or either partial or complete self-fertilization is the most frequent breeding system in endemic plant species (Cole, 2003), and it is generally associated with small, isolated and marginal populations that are visited relatively infrequently by pollinators. Such a breeding system provides reproductive assurance for these populations (Martén-Rodríguez & Fenster, 2010). Further studies are needed to understand the breeding system in *P. tabacum*, the pollinators and their interactions with flowers of *P. tabacum*.

However, limited pollen flow and selfing should not result in the domination of one genotype in population T. Considering that a similar pollinator situation might occur in population T, the reason might be an historical bottleneck following colonization. Thus, self-fertilization in *P. tabacum* ensures that a single, fertile founder can produce offspring. Moreover, effective seed dispersal guarantees that offspring colonize new patches, causing one genotype to dominate the whole population.

Unlike population T, we found substantial genetic structure in population D. Given that, in general, cpDNA in angiosperms is inherited maternally and, in most species, it is dispersed only via seeds (McCauley, 1995), gene flow within *P. tabacum* populations should also be represented by the spatial distribution patterns of cpDNA haplotypes. Thus, both the spatial distribution patterns of cpDNA haplotypes and asymmetrical gene flow among the three entrances of population D suggest that the *P. tabacum* individuals in the three entrances should be considered carefully. The lower cave entrance deserves initial attention insofar as it includes most of the *P. tabacum* individuals in population D. However, this entrance is more likely to be an admixing centre than to serve as the origin for *P. tabacum* in population D. First, 90% of the individuals at this entrance had the H2 haplotype, although H2 was not common in subpopulation D2 and absent at the upper cave entrance. Second, gene flow analysis showed asymmetrical gene flow from the upper cave entrance and subpopulation D2 to this entrance (Table 3). Finally, we observed that two individuals in P9 at the lower cave entrance shared an identical microsatellite multilocus genotype and cpDNA haplotype H3 with four individuals in subpopulation D2 (Fig. 2). Given that the H3 haplotype was very scarce at the lower cave entrance (three individuals among 97 samples), and P9 is the innermost patch located within the cave, the dispersal of seeds from these two individuals uphill to subpopulation D2 is considered to be unlikely. In contrast, the presence of these two individuals in P9 is understandable if wind-blown self-fertilized seeds are dispersed from subpopulation D2 to the lower cave entrance, which is at a lower elevation than subpopulation D2.

The upper cave entrance also deserves attention. The cpDNA haplotype H2 was predominant in population D. Although TCS analysis indicated that H2, with the highest outgroup probability (0.5), was the ancestral haplotype, this seems unlikely. Haplotypes H1 and H2 only differed by an 8-bp insertion (I1, Table 4) in the $trnL^{UAA}$–$trnF^{GAA}$ region. Although not found in haplotype H2 and H3, this 8-bp insertion was found in H4, and in sequences of the same region obtained previously for *P. tabacum* (GenBank accession numbers AJ492300.1 and GQ497200.1) and its sister species *Chirita longgangensis* (Möller *et al*., 2009) (GenBank accession numbers AJ492290.1 and DQ872809.1). Therefore, it is unlikely that the 8-bp insertion is later derived. Therefore, on the basis of the TCS results, haplotype H1 is hypothesized to be ancestral, haplotype H2 to be derived directly from haplotype H1, and haplotype H3 to be derived from haplotype H2 in population D. Given that plants growing at the upper cave entrance only carry haplotype H1, the results of the historical asymmetric gene flow analysis (Table 3) corroborate the proposal that the upper cave entrance is an ancestral location for *P. tabacum*. However, considerably fewer individuals grow at the upper cave entrance than at the lower entrance. An explanation for this unexpected result might be that, approximately 6 years ago, a survey at the site recorded many *P. tabacum* individuals growing at the upper cave entrance (S.-J. Li, pers. comm..). Subsequently, to attract tourists, the tourism company decorated the upper cave entrance and used sulphuric acid to wash the cave rock face to remove plants, which must have reduced greatly the number of *P. tabacum* individuals now growing at the upper cave entrance.

Owing to the uniqueness of *P. tabacum* habitats (Ren *et al*., 2010b) and their continuous deterioration caused by deforestation and economic developments in China, there are now only three populations, and one of these is extremely small. Given that these populations are geographically discontinuous and that we observed only limited gene flow within and between *P. tabacum* populations, all of these three

remnant populations must be highly isolated genetically. Therefore, from a conservation perspective, they all need protection to ensure maximal preservation of their genetic diversities and avoidance of extinctions.

The analysis of the two large populations studied here suggests that the structured gene pools within and among *P. tabacum* populations will warrant separate attention even for overall protection. Population D should receive the highest conservation priority because it has greater genetic diversity than population T. Protection of a small number of individuals within population T should be adequate to preserve the genetic diversity of the entire population. Within population D, although the population at the lower cave entrance might be admixed from that at the upper cave entrance and subpopulation D2, it remains the main distribution centre, and contains the highest level of genetic diversity. It should be given the highest conservation priority. Next, because individuals located at the upper cave entrance are indicated to be ancestral genetically and to play the most important role in the migration patterns of population D, conservation of individuals at the upper cave entrance should be prioritized over that of subpopulation D2. In addition, given that population D is located within a highly disturbed area, the long-term conservation of *P. tabacum* should include the re-establishment of new, genetically diverse populations at safe and undisturbed locations.

ACKNOWLEDGEMENTS

We thank many employees of the Lianzhou Ideal Tourism Development Co. Ltd. and Lianzhou Tianxin forest farm nature reserve for assistance with field sampling. We also thank three anonymous reviewers for their helpful comments. This study was funded by the National Basic Research Program of China (973 Program, 2007CB4116) and the Science and Technology Project of Guangdong Province, China (2011B060400003).

REFERENCES

Avise JC, Selander RK. 1972. Evolutionary genetics of cave-dwelling fishes of the genus *Astyanax*. *Evolution* **26:** 1–19.

Beerli P, Felsenstein J. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* **152:** 763–773.

Beerli P, Felsenstein J. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America* **98:** 4563–4568.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F. 1996–2004. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier.

Bradic M, Beerli P, León FJG, Esquivel-Bobadilla S, Borowsky RL. 2012. Gene flow and population structure in the Mexican blind cavefish complex (*Astyanax mexicanus*). *BMC Evolutionary Biology* **12:** 9.

Christman MC, Culver DC, Madden M, White D. 2005. Patterns of endemism of the eastern North American cave fauna. *Journal of Biogeography* **32:** 1441–1452.

Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9:** 1657–1660.

Cole CT. 2003. Genetic variation in rare and common plants. *Annual Review of Ecology, Evolution, and Systematics* **34:** 213–237.

Demesure B, Sodzi N, Petit RJ. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology* **4:** 129–131.

Doyle JJ. 1991. DNA protocols for plants – CTAB total DNA isolation. In: Hewitt GM, Johnston A, eds. *Molecular techniques in taxonomy*. Berlin: Springer-Verlag, 283–293.

Ebert D, Peakall R. 2009. A new set of universal *de novo* sequencing primers for extensive coverage of noncoding chloroplast DNA: new opportunities for phylogenetic studies and cpSSR discovery. *Molecular Ecology* **9:** 777–783.

Escudero A, Iriondo JM, Torres ME. 2003. Spatial analysis of genetic diversity as a tool for plant conservation. *Biological Conservation* **113:** 351–365.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14:** 2611–2620.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164:** 1567–1587.

Flot J-F, Wörheide G, Dattagupta S. 2010. Unsuspected diversity of *Niphargus amphipods* in the chemoautotrophic cave ecosystem of Frasassi, central Italy. *BMC Evolutionary Biology* **10:** 171.

Glime JM. 2007. *Bryophyte ecology*. Volume 1. Physiological ecology. Ebook sponsored by Michigan Technological University and the International Association of Bryologists. [Cited 27 March 2012]. Available at: http://www.bryoecol.mtu.edu/

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41:** 95–98.

Hance D. 1883. *Primulina tabacum* Hance. *Journal of Botany, British and Foreign* **21:** 169–170.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23:** 1801–1806.

Li SJ, Zhang DX, Zhang RQ, Cao M, Luo SX. 2006. Self-pollination by corolla-shedding in an endangered species *Primulina tabacum* (Gesneriaceae). In: Plant Systematics and Evolution Section of Botanical Society of China, ed. *Proceedings of the 9th Systematic and Evolution Symposium in China. Xi'an, China*. Unpublished, 41.

Martén-Rodríguez S, Fenster CB. 2010. Pollen limitation and reproductive assurance in Antillean Gesnerieae: a specialist vs. generalist comparison. *Ecology* **91:** 155–165.

McCauley DE. 1995. The use of chloroplast DNA polymorphism in studies of gene flow in plants. *Trends in Ecology & Evolution* **10:** 198–202.

Möller M, Pfosser M, Jang C-G, Mayer V, Clark A, Hollingsworth ML, Barfuss MHJ, Wang Y-Z, Kiehn M, Weber A. 2009. A preliminary phylogeny of the 'Didymocarpoid Gesneriaceae' based on the three molecular data sets: incongruence with available tribal classifications. *American Journal of Botany* **96:** 989–1010.

Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89:** 583–590.

Ni X, Huang Y, Wu L, Zhou R, Deng S, Wu D, Wang B, Su G, Tang T, Shi S. 2006. Genetic diversity of the endangered Chinese endemic herb *Primulina tabacum* (Gesneriaceae) revealed by amplified fragment length polymorphism (AFLP). *Genetica* **127:** 177–183.

Nybom H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology* **13:** 1143–1155.

Peakall R, Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6:** 288–295.

Peng SL, Cheng WC. 2002. *Rare and endangered plants in Guangdong*. Beijing: Science Press.

Pereira MJR, Salgueiro P, Rodrigues L, Coelho MM, Palmeirim JM. 2009. Population structure of a cave-dwelling bat, *Miniopterus schreibersii*: does it reflect history and social organization? *Journal of Heredity* **100:** 533–544.

Pipan T, Culver D. 2012. Convergence and divergence in the subterranean realm: a reassessment. *Biological Journal of the Linnean Society* **107:** 1–14.

Pritchard JK, Stephens M, Donnelly YP. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

Ren H, Ma G, Zhang Q, Guo Q, Wang J, Wang Z. 2010a. Moss is a key nurse plant for reintroduction of the endangered herb, *Primulina tabacum* Hance. *Plant Ecology* **209:** 313–320.

Ren H, Zhang Q, Wang Z, Guo Q, Wang J, Liu N, Liang K. 2010b. Community ecology knowledge for conservation through possible re-introduction: a case study of a rare plant *Primulina tabacum* Hance in China. *Plant Species Biology* **25:** 43–50.

Renno JF, Gazel C, Miranda G, Pouilly M, Berrebi P. 2007. Delimiting species by reproductive isolation: the genetic structure of epigean and hypogean *Trichomycterus* spp. (*Teleostei, Siluriformes*) in the restricted area of Torotoro (Up Amazon, Bolivia). *Genetica* **131:** 325–336.

Rousset F. 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* **8:** 103–106.

Schilthuizen M, Rutten EMJ, Haase M. 2012. Small-scale genetic structuring in a tropical cave snail and admixture with its above-ground sister species. *Biological Journal of the Linnean Society* **105:** 727–740.

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* **92:** 142–166.

Taberlet P, Gielly L, Pautou G, Bouvet J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17:** 1105–1109.

Wang Z-F, Ren H, Li Z-C, Zhang Q-M, Liang K-M, Ye W-H, Wang Z-M. 2013. Data from: Local genetic structure in the critically endangered, cave-associated perennial herb *Primulina tabacum* (Gesneriaceae). *Dryad Digital Repository*. doi:10.5061/dryad.nv066.

Wang Z-F, Ren H, Zhang Q-M, Ye W-H, Liang K-M, Li Z-C. 2009. Isolation and characterization of microsatellite markers for *Primulina tabacum*, a critically endangered perennial herb. *Conservation Genetics* **10:** 1433–1435.

Wilson GA, Rannala B. 2003. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163:** 1177–1191.

ARCHIVED DATA

Micosatellite genotyping data have been deposited at Dryad (Wang *et al.*, 2013).

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. Location of sampled patches of *Primulina tabacum* individuals in population D in Lianzhou City, Guangdong province, China. Subpopulation D1: upper cave entrance and lower cave entrance.