

Extensive population expansion of *Pedicularis longiflora* (Orobanchaceae) on the Qinghai-Tibetan Plateau and its correlation with the Quaternary climate change

FU-SHENG YANG,* YU-FEI LI,* XIN DING* and XIAO-QUAN WANG*†

*State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, 20 Nanxincun, Xiangshan, Beijing 100093, China, †Graduate University of the Chinese Academy of Sciences, Beijing 100039, China

Abstract

The Qinghai-Tibetan Plateau (QTP) is thought to be more strongly affected by the Quaternary glaciations than most other regions of the same latitude. It would be of great interest to investigate the population genetic structure of organisms distributed on the platform and its correlation with the Quaternary climatic oscillations. Here we used the chloroplast (cp)DNA *trnT-trnF* sequence to study genetic variation and phylogeography of *Pedicularis longiflora*, an alpine herb with extensive distribution on the QTP. Based on a range-wide sampling comprising 41 populations and 910 individuals, we detected 30 cpDNA haplotypes that were divided into five clades by phylogenetic and network analyses and a strong phylogeographical structure. All haplotypes but one in the three basal clades occur exclusively in the southeast QTP, whereas haplotypes in the young clade V occupy almost the whole species range. In particular, the young haplotype H18 occurs in 420 individuals, even at a frequency of 100% in some QTP platform populations and the Altai population. The haplotype distribution pattern, together with molecular clock estimation and mismatch distribution analysis, suggests that the southeast QTP was either a refuge for *P. longiflora* during the Quaternary climatic change or is the place of origin of the species. The present wide distribution of the species on the QTP platform has resulted from recent population expansions which could be dated back to 120 000–17 000 years ago, a period mostly before the last glacial maximum. The possible relationships among geographic genetic structure, climatic change and species diversification in *Pedicularis* are also discussed.

Keywords: colonization, cpDNA *trnT-trnF*, Himalayan-Hengduan Mountains, *Pedicularis longiflora*, phylogeography, Quaternary glacier

Received 18 May 2008; revision received 11 August 2008; accepted 17 September 2008

Introduction

The Qinghai-Tibetan Plateau (QTP), occupying approximately 2.5×10^6 km² or one-quarter of China (Zhang *et al.* 2002), is the largest and highest plateau in the world with an average altitude of 4500 m above sea level. In the early Tertiary, the Indian Plate collided with the Eurasian Plate, directly resulting in the subsequent uplift of the QTP (Zheng & Yao 2004). The biota of the QTP and its neighbouring mountains has changed dramatically due to the alteration of topography and local climate. It is more and more clear

that extant plants and animals on the plateau are colonists from other areas or endemic species derived recently (e.g. Wu 1980; Yang *et al.* 2003; Wei & Wang 2004; Liu *et al.* 2006; Peng *et al.* 2006; Ran *et al.* 2006; Qiao *et al.* 2007). However, the development of their modern distribution, i.e. how they colonized the plateau platform and expanded their distribution to the current range, is still in controversy. On the other hand, the range and intensity of the Quaternary glacier on the QTP have been hotly debated and the debate is still far from being resolved (Lehmkuhl & Owen 2005; Ehlers & Gibbard 2007). In the absence of sufficient fossil and palaeo-ecological evidence, variation in the distribution of organisms on the QTP in response to the Quaternary climatic changes is poorly understood. Fortunately, the phylogeographic method, since it was put forward twenty

Correspondence: Dr Xiao-Quan Wang, Fax: 86-10-62590843; E-mail: xiaoq_wang@ibcas.ac.cn (Fu-Sheng Yang, Yu-Fei Li contributed equally)

years ago, has proved to be very efficient in retrieving historical events of species, such as range fragmentation, refugial isolation, colonization and range expansion (Avice 2000).

To date, the Quaternary evolutionary histories of European and Northern American flora have been outlined by increasing phylogeographical studies (Petit *et al.* 1997; Abbott *et al.* 2000; Abbott & Brochmann 2003; Hewitt 2004; Petit *et al.* 2005; Anderson *et al.* 2006; Soltis *et al.* 2006). In contrast, in the QTP and its adjacent areas that span across the Indo-Burma and Himalaya biodiversity 'hotspots' (Myers *et al.* 2000; Biodiversity-hotspots 2005), only a few plant species have been phylogeographically studied (e.g. Song *et al.* 2003; Zhang *et al.* 2005; Gao *et al.* 2007; Meng *et al.* 2007; Yuan *et al.* 2008). Furthermore, all previous studies only sampled populations from the eastern edge of the QTP, and nearly all of them investigated tree species, such as conifers *Picea crassifolia* (Meng *et al.* 2007), *Pinus densata* (Song *et al.* 2003) and *Taxus wallichiana* (Gao *et al.* 2007). Interestingly, for most of the studied plants, only a single or very few haplotypes of maternally inherited cytoplasmic DNA were found in the populations located at the eastern edge of the QTP, indicating strong founder effects in the process of population expansion (Song *et al.* 2003; Zhang *et al.* 2005; Gao *et al.* 2007; Meng *et al.* 2007; Chen *et al.* 2008). It would be of great interest to know the genetic structure of populations in the main part or 'platform' of the QTP. An extensive investigation on the evolutionary history of species with a wide distribution in the QTP, especially in its central part, could not only retrieve the process of population establishment on the platform, but might also be helpful to explore the correlation between population genetic structure and the range and intensity of glaciations in the region. Recently, some phylogeographical studies have analysed some populations of birds and mammals from the QTP platform (Guo *et al.* 2006; Yang *et al.* 2006; Liu *et al.* 2007), but it is still very difficult to reveal evolutionary histories of the populations and to explore effects of the Quaternary glaciations on distribution and genetic structure of the platform populations due to strong mobility of the studied animals.

Pedicularis longiflora is one of the most widely distributed and ecologically significant plant species on the QTP, occurring throughout the Himalayan-Hengduan Mountains, and westward to the adjacent area such as Kashmir and probably the Pamirs (Dhar & Kachroo 1983; Shishkin & Bobrov 1994). Few sparse populations of the species are also present in the Altai and neighbouring mountains. However, this species does not occur in the vast area between the northern Tibetan Plateau and the Altai. Characterized by rosulate leaves and a special long corolla tube, this alpine plant usually grows in wet meadows or along hill-streams at elevations between 2600 and 5300 m (Yang *et al.* 1998). As an outcrossing and insect-pollinated dwarf herb, *P. longiflora* disperses its seeds largely by gravity. This herbaceous

species is ideal for studying the effects of Quaternary climatic oscillation on the genetic structure of plants on the QTP, since short-lived herbs, as opposed to trees, could have responded more quickly to environmental change on the Quaternary timescales (Comes & Kadereit 1998). Furthermore, previous molecular phylogenetic studies indicate that this species and/or its close relatives consist of a clade at a relatively basal position (Yang *et al.* 2003; Ree 2005; Yang & Wang 2007), and thus the phylogeographical study of *P. longiflora* would be helpful in interpreting the process of diversification and speciation of *Pedicularis*, a genus including over 500 species, half of which are endemic to the Himalayan-Hengduan Mountains.

In the present study, we used a nucleotide sequence of the chloroplast DNA *trnT-trnF* region, a marker widely used in plant phylogeographical analysis (e.g. Fujii *et al.* 1997; Zhang *et al.* 2005), to explore the genetic variation and to retrieve the phylogeographical history of *P. longiflora* based on a range-wide sampling. Then, we discussed the effects of environmental and climatic changes in the Quaternary era, such as glaciation, on distribution and evolution of plants on the QTP.

Materials and methods

Population sampling

During the summers of 2005 through 2007, population sampling was conducted throughout the range of *P. longiflora*. Fresh leaves were collected from 41 populations and, with few exceptions, 20–30 individuals that were at least 50 m apart from each other were sampled from each population (Table 1, Fig. 1). Only one population with 10 individuals was sampled from the Altai region, where the range of *P. longiflora* has sharply contracted due to serious degradation of the grassland (Drs DA German and SV Smirnov, Altai State University, Barnaul, personal communication). The latitude, longitude and altitude of each collection location were measured using an eTrex Global Positioning System (Garmin). In total, 910 individuals were sampled and leaves were dried with silica gel. *Pedicularis chinensis* and *P. siphonantha* were chosen as outgroups based on the results of previous morphological analysis (Tsoong 1955) and molecular systematic study of the genus (Ree 2005).

DNA extraction, PCR amplification and sequencing

Total genomic DNA was extracted from silica gel-dried leaves using the modified CTAB method (Rogers & Bendich 1988) and used as template in the polymerase chain reaction (PCR). The chloroplast (cp) DNA *trnT-trnF* region was initially amplified with primers 'a' and 'f' of Taberlet *et al.* (1991). Then two new primers specific to

Table 1 Geographic origins, sample sizes, cpDNA haplotypes and their frequencies of the 41 *Pedicularis longiflora* populations studied

Populations	Code	N	Longitude	Latitude	Altitude (m)	Haplotypes (Frequencies, %)	
1	Xianggelila, YN	XGLL	23	99°42'	27°43'	3315	H18(26), H23(48), H26(26)
2	Deqen, YN	BM	27	99°00'	28°22'	4347	H14(70), H18(4), H19(26)
3	Zayu, T	ZY	24	97°36'	28°48'	3620	H18(88), H28(13)
4	Muli, SC	ML	23	101°28'	28°04'	3550	H1(100)
5	Daocheng, SC	WM	30	100°03'	29°07'	4705	H2(30), H3(67), H26(3)
6	Litang, SC	HZL	26	100°11'	29°27'	4600	H3(88), H4(12)
7	Batang, SC	HZB	23	99°37'	30°16'	4470	H10(48), H13(9), H18(4), H26(35), H27(4)
8	Yajiang, SC	YJ	20	100°51'	30°00'	4215	H11(40), H18(30), H21(5), H26(5), 27H(20)
9	Litang, SC	LDS	24	100°15'	31°13'	4340	H12(33), H13(4), H18(4), H21(58)
10	Kangding, SC	KD	26	101°45'	30°03'	3860	H5(27), H6(23), H18(50)
11	Xiaojin, SC	XJ	26	102°19'	31°39'	3740	H8(100)
12	Barkam, SC	BK	28	102°31'	32°11'	3630	H8(18), H9(54), H16(25), H26(4)
13	Songpan, SC	SP	30	103°38'	32°46'	3250	H7(97), H18(3)
14	Hongyuan, SC	HY	25	102°22'	32°39'	3525	H7(44), H8(40), H13(4), H18(12)
15	Yushu 2, QH	YS2	14	96°34'	32°37'	3900	H13(7), H18(43), H24(29), H26(22)
16	Nangqen, QH	NQ	7	95°55'	32°48'	3840	H11(14), H18(43), H24(29), H26(14)
17	Yushu 1, QH	YS1	12	96°41'	32°53'	4500	H18(92), H24(8)
18	Yushu, QH	YS	27	96°32'	33°12'	4213	H13(4), H18(93), H26(4)
19	Chengduo, QH	CD	15	96°26'	33°12'	4500	H13(7), H18(93)
20	Tarlag, QH	TAR	27	99°37'	33°43'	4011	H18(41), H20(4), H26(56)
21	Madoi, QH	BYK	22	97°56'	34°24'	4323	H18(100)
22	Maqin, QH	MQ	22	100°17'	34°17'	3997	H18(71), H26(29)
23	Tongde, QH	TD	20	100°40'	35°15'	3208	H8(100)
24	Madoi, QH	HSX	23	98°51'	35°06'	4236	H18(87), H25(9), H26(4)
25	Datong, QH	DT	21	101°33'	37°06'	2704	H15(91), H18(10)
26	Qilian, QH	QL	26	100°13'	38°03'	3531	H15(69), H17(4), H18(27)
27	Damxung, T	DX	22	91°11'	30°34'	4278	H13(32), H18(64), H24(5)
28	Nyingchi, T	NC	23	94°25'	29°37'	3045	H18(87), H22(13)
29	Cuona, T	CN	27	91°55'	27°58'	4364	H18(63), H26(37)
30	Yadong, T	YD	30	88°58'	27°30'	3366	H18(100)
31	Dinggye, T	DY	21	87°47'	28°20'	4201	H18(100)
32	Namling, T	NML	14	89°02'	29°41'	4231	H18(93), H28(7)
33	Saga, T	SG	22	85°57'	29°27'	4700	H18(86), H28(14)
34	Nyalam, T	NLM	17	85°58'	28°08'	3790	H18(53), H28(47)
35	Gyirong, T	GY	20	85°18'	28°51'	4117	H18(80), H28(20)
36	Zhongba, T	ZB	24	84°13'	29°51'	4715	H18(54), H28(42), H30(4)
37	Coqen, T	CQ	23	85°03'	31°10'	4950	H18(100)
38	Burang, T	BR	22	81°16'	30°11'	3750	H28(96), H29(5)
39	Gar, T	GAR	23	79°43'	32°24'	4246	H18(100)
40	Rutog, T	RT	23	79°38'	33°25'	4271	H18(13), H28(87)
41	Altai, USSR	ALT	10	87°53'	49°17'	2600	H18(100)

Abbreviations: SC, Sichuan Province; YN, Yunnan Province; QH, Qinghai Province; T, Tibet Autonomous Region; N, number of sampled individuals.

Pedicularis (trna1: 5' CTAACCTCTGAGCTAAGCGGGC 3'; trnf1: 5' GGGTTACAAAGTTTTCTCTAAGTC 3') were designed based on the obtained sequences and used in all amplifications. The PCR mixture and amplification program followed Yang & Wang (2007). The PCR products were purified using the GFX PCR DNA and Gel Band Purification Kit (Pharmacia), and then sequenced with the two PCR primers (trna1, trnf1) and internal primers C (Taberlet *et al.* 1991) and I (5' CCATTTGTTAGAACAGCTTC 3'), using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction

Kit or DYEnamic Energy Transfer (ET) Terminator Reagent Premix Kit. After precipitation in 95% EtOH and 3M NaAc (pH 5.3), the sequencing products were separated on either an ABI PRISM 3730xl or MegaBACE 1000 DNA analyser.

Data analyses

The DNA sequences were aligned using the program CLUSTAL x (Thompson *et al.* 1997) and manually adjusted in BioEdit. All sequences have been deposited in GenBank

1000 replications was performed to quantify the confidence in the partitioning within the trees.

To detect historical population expansion events in *P. longiflora*, mismatch distributions were calculated using the program ARLEQUIN version 3.1 (Excoffier *et al.* 2005). A total of 1000 parametric bootstrap replicates were used to generate an expected distribution under a model of sudden demographic expansion (Rogers & Harpending 1992). The sum-of-squared deviations (SSD) between observed and expected mismatch distribution were computed and *P*-values were calculated as the proportion of simulations producing a larger SSD than the observed SSD. The raggedness index and its significance were also calculated to quantify the smoothness of the observed mismatch distribution. If the sudden expansion model was not rejected, the relationship $\tau = 2ut$ (Rogers & Harpending 1992) was used to estimate the age of expansion (*t*), where μ is the mutation rate for the DNA sequences.

To examine rate constancy among lineages, the likelihood ratio test (Felsenstein 1988) was carried out using PAUP* 4.10b (Swofford 2002), in which likelihood scores of the trees with and without an enforced molecular clock were compared. Significance was calculated by comparing two times the difference in log likelihoods to a χ^2 distribution with $n - 2$ degrees of freedom, where n is the number of taxa (haplotypes of *P. longiflora* plus outgroups). When the molecular clock hypothesis was not rejected, the average divergence times for different *P. longiflora* lineages were estimated from the uncorrected mean pairwise distance between sequences using the variation rate of 8.24×10^{-9} substitution per site per year (s/s/y) for the *trnL-trnF* region in herbs as summarized in Richardson *et al.* (2001), since no fossil record of Orobanchaceae is available to calibrate nucleotide substitution rate in *Pedicularis* (Lousewort). This mutation rate was also used to estimate the age of population expansion (*t*) based on the formula $\tau = 2ut$ with a generation time of one year being assumed for *P. longiflora*. For comparison, we also estimated the divergence time in *P. longiflora* using Bayesian inference under relaxed-clock models implemented by multivtime (Rutschmann 2005), in which posterior distributions were approximated by sampling values at every 100 steps over 1 000 000 Markov chain Monte Carlo (MCMC) steps after discarding a burn-in of 100 000 steps. The time of the basal node inferred from the average evolutionary rate was used as an age constraint.

Results

Intraspecific sequence variation and population genetic structure

The cpDNA *trnT-trnF* region of *Pedicularis longiflora* populations ranged from 1444 bp to 1475 bp in size. The sequence

alignment of ingroups had 1499 bp, including 45 polymorphic sites, of which 41 were parsimony-informative. These polymorphic sites defined 30 haplotypes (H1–H30) that were asymmetrically distributed across the species range (Table 1, Fig. 1). The 14 populations from the southeastern QTP, occupying a small portion of the distribution, harboured 22 haplotypes (73% of the total), in which 16 (H1–H7, H9, H10, H12, H14, H16, H19, H21, H23, H27) were endemics. In contrast, the remaining 27 populations, representing the main range of the species (the QTP platform), had only 14 haplotypes, and eight (H15, H17, H20, H22, H24, H25, H29, H30) of them were endemic. Total haplotype diversity (h_T) was estimated to be 0.770 and within-population diversity (h_S) was 0.332. Significant population differentiation was observed, with a $G_{ST} = 0.564$ and a $N_{ST} = 0.812$, and the permutation test showed that G_{ST} and N_{ST} were significantly different from each other ($N_{ST} > G_{ST}$, $U = 3.27$, $P < 0.01$). The spatial analysis of molecular variance (SAMOVA) indicated that the genetic differentiation among groups was the highest ($F_{CT} = 0.8428$, $P < 0.001$) when samples were pooled into two groups, one including three populations (4, 5, 6) from the southeastern-most QTP (SEM-QTP) (Fig. 1), and the other comprising the remaining 38 populations (R-QTP). The geographic ranges of the two groups were non-overlapping. The AMOVA revealed that 84.28% of the total variation occurred among groups, 10.25% among populations, and 5.47% within populations. The pairwise F_{ST} value between the two groups was 0.833.

Phylogenetic relationships and distributions of the cpDNA haplotypes

Topology of the NJ tree of the 30 haplotypes detected from the 910 individuals of *P. longiflora* is shown in Fig. 2, which was the same as the MP and ML trees in the major clades. When more outgroups were added, the tree topology (not shown) remained unchanged. Monophyly of the 30 haplotypes was strongly supported ($\geq 99\%$ bootstrap support) and five clades were recognized. Haplotypes in the three basal clades (I, II, III) occurred only in populations from the southeast edge of the QTP except that H8 was also found in population 23 (Fig. 1). All the four haplotypes (H1–H4) in Clade I were confined to the three populations (4, 5, 6) that formed the SEM-QTP regional group recognized by the SAMOVA analysis. Clade II haplotypes (H5, H6) were endemic to population 10, while Clade III haplotypes (H7–H10) were distributed in the northern populations except H10 from population 7. The evolutionary relationships of haplotypes in the two sister clades IV and V were poorly resolved, showing polytomies that might have stemmed from rapid radiations. Clade IV included seven haplotypes that were sparsely distributed in some populations from the eastern QTP, while Clade V haplotypes were very widely distributed in 37 out of the total 41 populations. Of

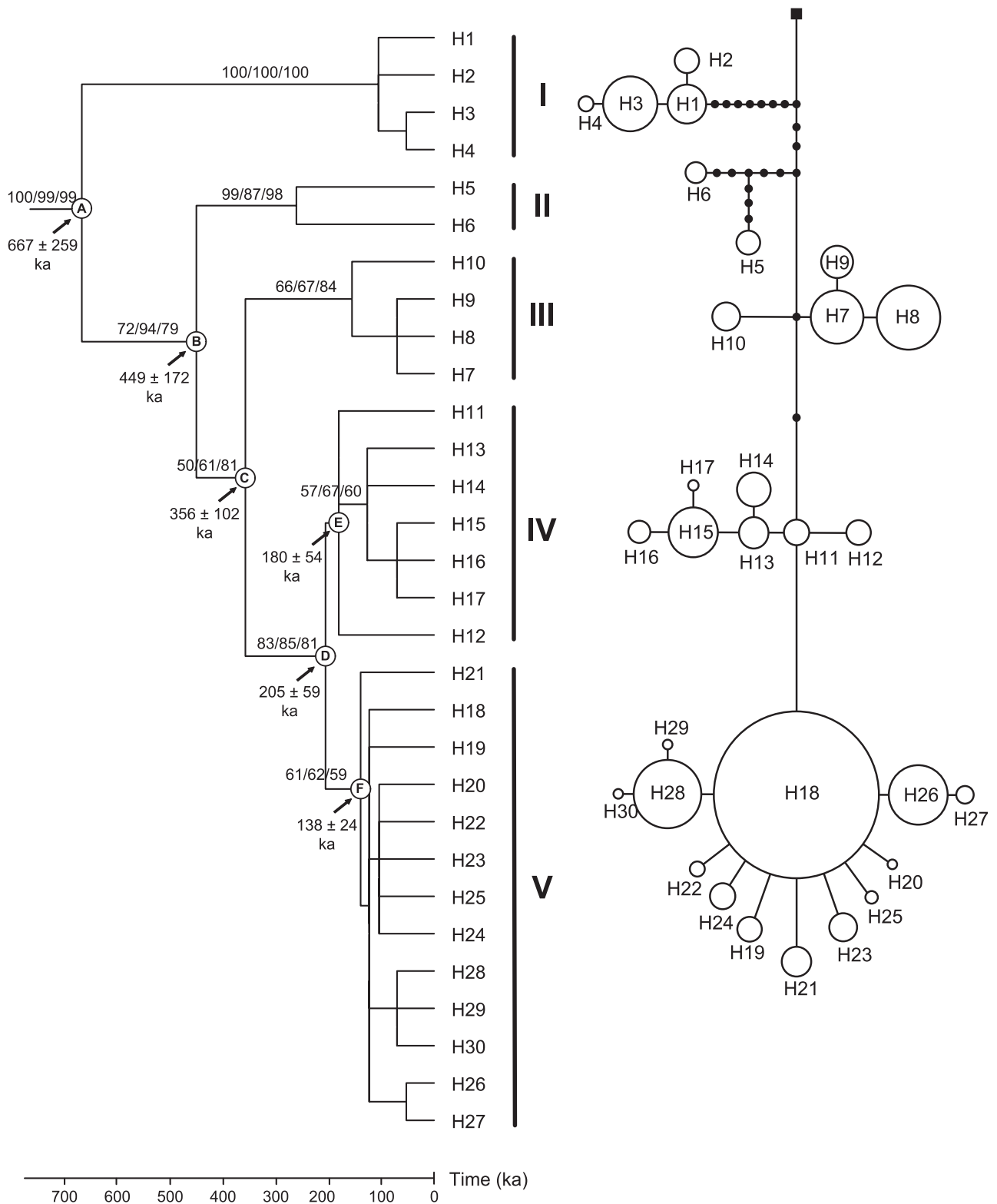


Fig. 2 The NJ tree topology (Left) and network (Right) of the 30 cpDNA haplotypes detected from the *trnT-trnF* region of *Pedicularis longiflora* and their divergence times estimated with the average evolutionary rate. The node age estimates with stand errors are marked under branches. Numbers on the branches indicate the bootstrap values for NJ (Left, 1000 replicates), MP (Middle, 1000 replicates) and ML (Right, 500 replicates) analyses, respectively. The relative sizes of the circles in the network are proportional to haplotype frequencies and black dots represent missing haplotypes (not sampled or extinct).

Table 2 Results of the mismatch distribution analysis for the 32 *Pedicularis longiflora* populations with intra-population divergence

Populations	N	Tau	Age (ka)	SSD P-value	Raggedness index
1 XGLL	23	1.146	46	NS	0.093
2 BM	27	0	—	0.001	0.530
3 ZY	24	2.930	119	NS	0.347
5 WM	30	2.875	116	NS	0.621
6 HZL	26	2.930	119	NS	0.362
7 HZB	23	5.807	235	NS	0.221
8 YJ	20	1.467	59	NS	0.057
9 LDS	24	3.850	—	NS	0.509*
10 KD	26	9.572	—	NS	0.294*
12 BK	28	8.127	—	NS	0.447*
13 SP	30	3.000	121	NS	0.880
14 HY	25	6.078	246	NS	0.167
15 YS2	14	1.203	49	NS	0.129
16 NQ	7	1.336	54	NS	0.313
17 YS1	12	2.965	120	NS	0.472
18 YS	27	3.000	121	NS	0.618
19 CD	15	3.215	130	NS	0.286
20 TAR	27	0.766	—	NS	0.251*
22 MQ	22	0.623	25	NS	0.204
24 HSX	23	3.000	121	NS	0.320
25 DT	21	3.000	—	0.014	0.736
26 QL	26	0	—	< 0.001	0.518
27 DX	22	1.875	—	NS	0.468*
28 NC	23	2.930	119	NS	0.332
29 CN	27	0.730	30	NS	0.236
32 NML	14	0.207	8	NS	0.531
33 SG	22	2.930	119	NS	0.317
34 NLM	17	0.814	33	NS	0.284
35 GY	20	0.453	17	NS	0.219
36 ZB	24	0.783	32	NS	0.218
38 BR	22	3.000	121	NS	0.677
40 RT	23	2.930	119	NS	0.333

* $P < 0.05$; NS: not significant. Population numbers are the same as those in Table 1.

the 13 haplotypes (H18–H30) in Clade V, H18 was the most dominant haplotype, occurring in 420 individuals of 34 populations, especially with very high frequencies (even 100%) in the westward populations from the QTP platform. However, most of the other haplotypes were detected only in one or two populations with low frequencies. According to the network of cpDNA haplotypes (Fig. 2), H18 was also the ancestral haplotype that gave rise to the other haplotypes in Clade V. It is interesting that the remote northern Altai population has pure H18.

Mismatch distribution analysis

The mismatch distribution analysis indicated that 24 of the 32 populations with intra-population divergence had SSD and raggedness index values that did not reject a sudden expansion model. These populations had expansion ages

Table 3 Molecular clock estimation of divergence times (thousand years ago) for lineages in *Pedicularis longiflora* based on the cpDNA *trnT-trnF* region (95% credibility intervals in parentheses). Nodes correspond to those marked on Fig. 2

Node	Method	
	Average sequence diversity	Bayesian
A	667 (± 259)	—
B	449 (± 172)	471 (± 107)
C	356 (± 102)	322 (± 110)
D	205 (± 59)	198 (± 95)
E	180 (± 54)	198 (± 95)
F	138 (± 24)	99 (± 74)

that ranged from 246 to 8 thousand years ago (Table 2), including two earlier than 200 thousand years ago (HY, HZB), 21 within 130 to 17 thousand years ago, and one later than 10 thousand years ago (NML).

Molecular clock test and divergence time dating

The result of the LRT test did not reject a clock-like evolution for the cpDNA haplotypes ($\delta = 23.17$, d.f. = 30, $P > 0.05$), so the previously reported divergence rate of the *trnL-trnF* region for herbaceous plants (Richardson *et al.* 2001) was used to estimate divergence times of the *P. longiflora* lineages (haplotypes), which were indicated on the tree in Fig. 2. These divergence times, ranging from 667 \pm 259 thousand years ago at node A to 138 \pm 24 thousand years ago at node F, were largely consistent with the ages estimated using the Bayesian method (Table 3).

Discussion

Extensive population expansion of *P. longiflora* throughout the QTP platform in the Quaternary

In the present study, 30 cpDNA haplotypes were detected from *Pedicularis longiflora* based on a range-wide sampling ($n = 910$), and divergence times for all the haplotypes were estimated to be less than 1 million years (Table 3). Haplotypes in the three basal clades (I, II, III) and clade IV are endemic to the southeastern or eastern edge of the QTP except that some haplotypes from clade IV, such as H11 and H13, occur in several populations on the eastern platform of the plateau at low frequencies. The 13 haplotypes comprising H18 and its offspring in the young Clade V (138 \pm 24 thousand years ago), sister to clade IV, cover 37 out of the 41 populations we studied, but most of them are only present in one or two populations with low percentages, respectively. It is surprising that H18 occurs in 420 individuals of 34 populations, especially in the overwhelming majority of individuals from the QTP platform, even being fixed in

the westward populations at a frequency of 100% (Table 1, Figs 1 and 2). This haplotype distribution pattern and molecular clock estimation suggest that the southeastern or eastern edge of the QTP could have served as refuge for *P. longiflora* during the Quaternary climatic change. The alternative explanation is that this species originates in the southeastern QTP, given that its close relatives also occur in this area (Tsoong 1955; Ree 2005). After the origin of *P. longiflora*, rapid intraspecific diversification could have been triggered by the Quaternary climatic oscillations. The present wide distribution of the species in the QTP platform, especially in the central and western part, would have resulted from recent population expansions, similar to the phenomenon of rapid recolonization that has been reported from many species of temperate Europe and North America (Hewitt 1996, 2004). When an area was swept by the Quaternary glaciations and recolonized rapidly by leading edge populations from borders of refugia, it would be much more difficult for those behind the pioneers to advance (Hewitt 2000), producing reduced allele diversity and areas of genetic homogeneity.

It is very likely that one or more leading populations carrying some clade V haplotypes, in particular the most dominant haplotype H18, had rapidly colonized or recolonized the QTP platform during interglacial periods, and bottleneck or founder effects would be responsible for genetic homogeneity in the platform populations, especially the western populations (Fig. 1). One may argue that this haplotype distribution pattern could be the result of colonization of a few haplotypes from the diversification centre (southeastern QTP) to the QTP platform after the last glacial maximum (LGM). However, by using the average substitution rate (8.24×10^{-9} s/s/y) of the *trnL-trnF* region, the age of *P. longiflora* was estimated about 3.47 (± 0.44) million years (calculation details not shown), and expansion times estimated for most populations of the species (Table 2) are also much earlier than the LGM. Nevertheless, this scenario could not be ruled out in consideration of the complicated process of the Quaternary glaciations in the QTP and the accuracy of the molecular clock estimation, although evolutionary rate accelerations in parasitic plants are lineage-specific and there is no significant increase in the rate of plastid genes in *Pedicularis* (dePamphilis *et al.* 1997; Young & dePamphilis 2005; Yang & Wang 2007). The fact that the Altai population harbours pure H18 also indicates a rapid process of population colonization throughout the QTP and northward to Altai, accompanied with strong founder effects. The Altai population could not have originated from a long-distance dispersal after the development of the wide arid belt in northern China and Republic of Mongolia, considering the low dispersal ability of Lousewort seed (Tsoong 1955). In addition, the colonization to the QTP platform likely has occurred more than once since some haplotypes in clade IV such as H11 and H13 are

also distributed in the eastern part of the platform, although these haplotypes could have been transferred to the platform by seed gene flow.

The inference of a past rapid colonization of *P. longiflora* on the QTP is also supported by the mismatch distribution analysis, in which 24 of the 32 populations with intra-population diversity do not reject a sudden expansion expectation (Table 2). For all but one (population NML) of the 16 populations on the platform with an inferred sudden expansion history, the estimated times of population expansion are 120–17 thousand years ago (Table 2), mostly corresponding to the last interglacial period in the late Pleistocene (Zheng *et al.* 2002). This age estimation is corroborated by the disjunct distribution of H18 between the QTP and the Altai Mountains, two areas separated by a wide arid belt that had mostly developed before the LGM (21–17 thousand years ago) (Yang *et al.* 2004). The sudden population expansion of *P. longiflora* on the platform after the penultimate glacial period (300–130 thousand years ago), most likely before LGM, is further supported by results of the molecular clock analysis, which indicates a burst of haplotype diversification in the clade V after 138 ± 24 thousand years ago, almost synchronous with expansion of the populations.

Actually, increasing studies have indicated that glaciations on the QTP have become less extensive throughout the Quaternary after the most extensive Naynayxungla Glaciation at 720–500 thousand years ago (Zheng *et al.* 2002; Lehmkuhl & Owen 2005; Owen & Benn 2005; Ehlers & Gibbard 2007), in particular after 170 thousand years ago (Schafer *et al.* 2002), although the Quaternary glaciations are asynchronous and have occurred to variable extents in different regions of the QTP (Owen *et al.* 2005; Thompson *et al.* 2006). The Quaternary Glacial Distribution Map of Tibet Plateau shows limited glaciation in the interior of the Tibetan Plateau but expanded ice caps and valley glaciers on its margins during the last glacial cycle (Li *et al.* 1991). Based on cosmogenic surface exposure ages of erratics on top of moraines, Schafer *et al.* (2002) found that glacial advances were restricted to a few tens of kilometres during the last 170 thousand years in Central Tibet and during the peak of the last glaciation in eastern Tibet, and advances of Tibetan glaciers were much less prominent than elsewhere in the northern hemisphere most likely due to very arid conditions and high sublimation rates. Recent studies also suggest limited glacier advances during the LGM (MIS-2) throughout the QTP (summarized in Lehmkuhl & Owen 2005). The ages (≤ 130 thousand years ago) of *P. longiflora* population expansion on the QTP platform we estimated do not conflict with the results of phylogeographical analysis on snow finch endemic to the platform (Qu *et al.* 2005) and are also comparable to the expansion times of terrestrial species with Holarctic ranges summarized by Hewitt (2004), but are earlier than the estimated expansion times

(8–3 thousand years ago) for *Juniperus przewalskii* and *Picea crassifolia*, two conifers with distributions extending only to the northeast edge of the QTP (Zhang *et al.* 2005; Meng *et al.* 2007).

Phylogeographic structure of P. longiflora and its correlation with the late Cenozoic climate change

During the Quaternary, climatic oscillations resulted in repeated drastic environmental changes, which further caused massive range shifts of most plants and animals, leading to population genetic consequences and particular phylogeographic patterns (Hewitt 2000). In *P. longiflora*, a strong phylogeographic structure was found between the QTP platform and the southeast QTP, which could also be correlated with the late Cenozoic climate change. The AMOVA analysis indicates that most of the haplotype variation (84.28%) occurs between the two geographic groups R-QTP and SEM-QTP, and a significant geographical structure ($N_{ST} > G_{ST}$, $P < 0.01$) was detected by the software HAPLONST. Moreover, phylogenetic analysis of the cpDNA haplotypes shows that the ancestral lineages (clades I, II, III) are almost exclusively distributed in the southeast QTP (Figs 1 and 2), a restricted area that harbours 73% of the total haplotypes. The rich haplotypes in the southeast QTP imply that the distribution of *P. longiflora* could have experienced several contraction/expansion cycles in response to the repeated advance/retreat of glaciations. When northern populations migrated to the southeast edge of the species range during the glacial periods, more and more haplotypes accumulated there, although some haplotypes in the southeast QTP could have originated *in situ*. On the other hand, given the great altitudinal variability and complicated topography in the southeast QTP, repeated range fragmentation and re-expansion of *P. longiflora* with climatic oscillations would have stimulated intraspecific diversification and population genetic differentiation of the species in the region.

It is interesting that haplotypes in the three basal lineages (clades I, II, III) occur in different geographical areas of the southeast QTP (Figs 1 and 2). The four haplotypes in clade I (H1–H4) are confined to the three populations (4, 5, 6) from the SEM-QTP, and thus they would not have contributed to northward and/or westward colonization of *P. longiflora* during interglacial periods. Similarly, H5 and H6 in clade II are endemic to population 10. The haplotypes in clade III (H7–H10) are distributed in population 7 and five northern populations (11, 12, 13, 14, 23), among which only H8 has been expanded into the eastern edge of the QTP platform. The three lineages that have low haplotype diversity and non-overlapping distribution could have been isolated from each other for a relatively long time. Although six of the ten populations mentioned above harbour some young haplotypes from clades VI and V, this is very likely a result

of recent population expansion or seed gene flow. Long-term persistence of isolated populations has been a common phenomenon in the Mediterranean, where the evolutionary history of some forest species could be traced to the early Tertiary based on analyses of population genetic structure and phylogeographic patterns in refugia (summarized by Petit *et al.* 2005). In *P. longiflora*, the divergence of haplotypes in the basal clades endemic to the southeast QTP can date back to the Naynayxungla Glaciation period (720–500 thousand years ago) (Zheng *et al.* 2002), and therefore these lineages would have survived the later glaciations in the refugial region (Fig. 2). This inference is also supported by the fact that the estimated expansion times for some populations in the southeast QTP, such as HZB, are substantially earlier (> 200 thousand years ago) than those for the populations on the QTP platform (Table 2).

Geographic genetic structure and species diversification in Pedicularis

It is well recognized that the Quaternary climatic oscillations have made great influences on genetic diversification and speciation in the mid- to high-latitude regions of Europe and North America (Comes & Kadereit 1998; Hewitt 2004). A strong correlation between population differentiation and species richness has been reported from groups as diverse as *Begonia* (Hughes & Hollingsworth 2008), Coleoptera (Ribera *et al.* 2003) and vertebrates (Martin & McKay 2004), suggesting a great influence of strong population structure on speciation. The geographic genetic structure of *P. longiflora* resulted from population isolation in the southeast QTP and recent rapid population expansion on the QTP platform is also consistent with the species diversity distribution of *Pedicularis* in the plateau. This genus has many more species in the southeast QTP (the Hengduan Mountains) than in the inner QTP platform, indicating a recent recolonization of some species to the platform from surrounding regions, a consequence of redistribution of species responding to the Quaternary climatic changes. The high species diversity in the southeast QTP was previously thought to have originated through genetic isolation in the diverse habitats as a result of the rapid uplift of the QTP (Wu 1980; Axelrod *et al.* 1996), although the effect of diversified climate was not ignored. However, the nearly ubiquitous distribution of clade V haplotypes of *P. longiflora* throughout the QTP provides strong evidence that the effect of topography alone is not enough to produce high genetic differentiation. The effect could have been greatly strengthened by repeated shifts of species range in the southeast QTP, and thus resulted in a much more rapid accumulation of genetic diversity and topographical isolation that are important for speciation to occur, considering the short generation time of the herbs like *Pedicularis*. Meanwhile, the extremely diversified habitats and low dispersal ability

of *Pedicularis* may also be responsible for interspecific genetic isolation and accelerated speed of speciation in the genus.

Acknowledgements

We thank the two anonymous reviewers and the Subject Editor for their valuable suggestions and comments on the manuscript. We also thank Drs Sergey V. Smirnov, Dmitry A. German, Yujin Wang, Yousheng Chen, Shilong Chen and Wenli Chen, Mrs Zhiheng Yang and Lun Suo for their help in the sample collection. This study was supported by National Basic Research Program of China (Grant no. 2007CB411600), the Chinese Academy of Sciences (Grant No. kzcx2-yw-415) and the National Natural Science Foundation of China (Grant Nos. 30730010, 30400023).

References

- Abbott RJ, Brochmann C (2003) History and evolution of the arctic flora: in the footsteps of Eric Hult n. *Molecular Ecology*, **12**, 299–313.
- Abbott RJ, Smith LC, Milne RI, Crawford RMM, Wolff K, Balfour J (2000) Molecular analysis of plant migration and refugia in the Arctic. *Science*, **289**, 1343–1346.
- Anderson LL, Hu FS, Nelson DM, Petit RJ, Paige KN (2006) Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *Proceedings of the National Academy of Sciences, USA*, **103**, 12447–12450.
- Avise JC (2000) *Phylogeography: the History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts.
- Axelrod DI, Al-Shehbaz I, Raven PH (1996) History of the modern flora of China. In: *Floristic Characteristics and Diversity of East Asian Plants* (eds Zhang AL, Wu SG), pp. 43–45. Springer, New York.
- Biodiversity-hotspots (2005) Conservation International. (Available at: http://www.biodiversityhotspots.org/xp/Hotspots/hotspots_by_region/Pages/default.aspx).
- Chen S, Wu G, Zhang D *et al.* (2008) Potential refugium on the Qinghai-Tibet Plateau revealed by the chloroplast DNA phylogeography of the alpine species *Metagentiana striata* (Gentianaceae). *Botanical Journal of the Linnean Society*, **157**, 125–140.
- Comes HP, Kadereit JW (1998) The effect of Quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science*, **3**, 432–438.
- Dhar U, Kachroo P (1983) *Alpine Flora of Kashmir Himalaya*. Scientific Publishers, Jodhpur, India.
- Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, **11**, 2571–2581.
- Ehlers J, Gibbard PL (2007) The extent and chronology of Cenozoic Global Glaciation. *Quaternary International*, **164–165**, 6–20.
- Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.
- Felsenstein J (1988) Phylogenies from molecular sequences: inference and reliability. *Annual Review of Genetics*, **22**, 521–565.
- Fujii N, Ueda K, Watano Y, Shimizu T (1997) Intraspecific sequence variation of chloroplast DNA in *Pedicularis chamissonis* Steven (Scrophulariaceae) and geographic structuring of the Japanese ‘Alpine’ plants. *Journal of Plant Research*, **110**, 195–207.
- Gao LM, Moller M, Zhang XM *et al.* (2007) High variation and strong phylogeographic pattern among cpDNA haplotypes in *Taxus wallichiana* (Taxaceae) in China and North Vietnam. *Molecular Ecology*, **16**, 4684–4698.
- Guo S, Savolainen P, Su JP *et al.* (2006) Origin of mitochondrial DNA diversity of domestic yaks. *BMC Evolutionary Biology*, **22**, 73.
- Hewitt G (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.
- Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.
- Hewitt G (2004) Genetic consequences of climatic oscillations in the Quaternary. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **359**, 183–195.
- Hughes M, Hollingsworth PM (2008) Population genetic divergence corresponds with species-level biodiversity patterns in the large genus *Begonia*. *Molecular Ecology*, **17**, 2643–2651.
- Kimura M (1981) Estimation of evolutionary distances between homologous nucleotide sequences. *Proceedings of the National Academy of Sciences, USA*, **78**, 454–458.
- Lehmkuhl F, Owen LA (2005) Late Quaternary glacialiation of Tibet and the bordering mountains: a review. *Boreas*, **34**, 87–100.
- Li B, Li J, Cui Z *et al.* (1991) *Quaternary Glacial Distribution Map of Qinghai-Xizang (Tibet) Plateau*. Science Press (Map), Beijing, China.
- Liu Z, Ren B, Wei F *et al.* (2007) Phylogeography and population structure of the Yunnan snub-nosed monkey (*Rhinopithecus bieti*) inferred from mitochondrial control region DNA sequence analysis. *Molecular Ecology*, **16**, 3334–3349.
- Liu JQ, Wang YJ, Wang AL, Hideaki O, Abbott RJ (2006) Radiation and diversification within the *Ligularia–Cremanthodium–Parasenecio* complex (Asteraceae) triggered by uplift of the Qinghai-Tibetan Plateau. *Molecular Phylogenetics and Evolution*, **38**, 31–49.
- Martin PR, McKay JK (2004) Latitudinal variation in genetic divergence of populations and the potential for future speciation. *Evolution*, **58**, 938–945.
- Meng L, Yang R, Abbott RJ *et al.* (2007) Mitochondrial and chloroplast phylogeography of *Picea crassifolia* Kom. (Pinaceae) in the Qinghai-Tibetan Plateau and adjacent highlands. *Molecular Ecology*, **16**, 4128–4137.
- Myers N, Mittermeier R, Mittermeier C, Fonseca G, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.
- Nei M (1975) *Molecular Population Genetics and Evolution*. North-Holland, Amsterdam and New York.
- Owen LA, Benn DI (2005) Equilibrium-line altitudes of the Last Glacial Maximum for the Himalaya and Tibet: an assessment and evaluation of results. *Quaternary International*, **138–139**, 55–78.
- Owen LA, Finkel RC, Barnard PL *et al.* (2005) Climatic and topographic controls on the style and timing of Late Quaternary glacialiation throughout Tibet and the Himalaya defined by ¹⁰Be cosmogenic radionuclide surface exposure dating. *Quaternary Science Reviews*, **24**, 1391–1411.
- dePamphilis CW, Young ND, Wolfe AD (1997) Evolution of plastid gene *rps2* in a lineage of hemiparasitic and holoparasitic plants: many losses of photosynthesis and complex patterns of rate variation. *Proceedings of the National Academy of Sciences, USA*, **94**, 7367–7372.
- Peng Z, Ho SYW, Zhang Y, He S (2006) Uplift of the Tibetan plateau: Evidence from divergence times of glyptosternoid catfishes. *Molecular Phylogenetics and Evolution*, **39**, 568–572.

- Petit RJ, Pineau E, Demesure B *et al.* (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences, USA*, **94**, 9996–10001.
- Petit RJ, Hampe A, Cheddadi R (2005) Climate changes and tree phylogeography in the Mediterranean. *Taxon*, **54**, 877–885.
- Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics*, **144**, 1237–1245.
- Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.
- Qiao CY, Ran JH, Wang XQ (2007) Phylogeny and biogeography of *Cedrus* (Pinaceae) inferred from sequences of seven paternal chloroplast and maternal mitochondrial DNA regions. *Annals of Botany*, **100**, 573–580.
- Qu YH, Ericson PGP, Lei FM, Li SH (2005) Postglacial colonization of the Tibetan plateau inferred from the matrilineal genetic structure of the endemic red-necked snow finch, *Pyrgilauda ruficollis*. *Molecular Ecology*, **14**, 1767–1781.
- Ran JH, Wei XX, Wang XQ (2006) Molecular phylogeny and biogeography of *Picea* (Pinaceae): implications for phylogeographical studies using cytoplasmic haplotypes. *Molecular Phylogenetics and Evolution*, **41**, 405–419.
- Ree RH (2005) Phylogeny and the evolution of floral diversity in *Pedicularis* (Orobanchaceae). *International Journal of Plant Science*, **166**, 595–613.
- Ribera I, Foster GN, Voger AP (2003) Does habitat use explain large scale species richness patterns of aquatic beetles in Europe? *Ecography*, **26**, 145–152.
- Richardson JE, Pennington RT, Pennington TD, Hollingsworth PM (2001) Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science*, **293**, 2242–2245.
- Rogers SO, Bendich AJ (1988) Extraction of DNA from plant tissues. *Plant Molecular Biology Manual*, **A6**, 1–10.
- Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.
- Rutschmann F (2005) *Bayesian Molecular Dating Using PAML/Multidivtime. A Step-by-Step Manual*. University of Zurich, Switzerland. Available at <http://www.plant.ch>.
- Schafer JM, Tschudi S, Zhao Z *et al.* (2002) The limited influence of glaciations in Tibet on global climate over the past 170 000 yr. *Earth and Planetary Science Letters*, **194**, 287–297.
- Shishkin BK, Bobrov EG (1994) *Flora of the USSR* 22 (English translation), p. 775. Bishen Singh Mahendra Pal Singh and Koeltz Scientific Books, Dehra Dun, India.
- Soltis DE, Morris AB, Lachlan JM, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, **15**, 4261–4293.
- Song BH, Wang XQ, Wang XR, Ding KY, Hong DY (2003) Cytoplasmic composition in *Pinus densata* and population establishment of the diploid hybrid pine. *Molecular Ecology*, **12**, 2995–3001.
- Swofford DL (2002) *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4.0. Sinauer Associates, Sunderland, MA.
- Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three noncoding regions of chloroplast DNA. *Plant Molecular Biology*, **17**, 1105–1109.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**, 4876–4882.
- Thompson LG, Mosley-Thompson E, Davis ME *et al.* (2006) Ice core evidence for asynchronous glaciation on the Tibetan Plateau. *Quaternary International*, **154–155**, 3–10.
- Tsoong PC (1955) A new system for the genus *Pedicularis*. *Acta Phytotaxonomica Sinica*, **4**, 71–147.
- Wei XX, Wang XQ (2004) Recolonization and radiation in *Larix* (Pinaceae): evidence from nuclear ribosomal DNA paralogues. *Molecular Ecology*, **13**, 3115–3123.
- Wu CY (1980) *Vegetation of China*. Science Press, Beijing, China.
- Yang H, Holmgren N, Mill R (1998) *Pedicularis* L. In: *Flora of China* 18 (eds Wu Z, Raven P), pp. 97–209. Science Press, Beijing, China.
- Yang X, Rost KT, Lehmkuhl F, Zhou Z, Dodson J (2004) The evolution of dry lands in northern China and in the Republic of Mongolia since the Last Glacial Maximum. *Quaternary International*, **118–119**, 69–85.
- Yang FS, Wang XQ (2007) Extensive length variation in the cpDNA *trnT-trnF* region of parasitic *Pedicularis* and its phylogenetic implications. *Plant Systematics and Evolution*, **264**, 251–264.
- Yang FS, Wang XQ, Hong DY (2003) Unexpected high divergence in nrDNA ITS and extensive parallelism in floral morphology of *Pedicularis* (Orobanchaceae). *Plant Systematics and Evolution*, **240**, 91–105.
- Yang S, Yin Z, Ma X, Lei F (2006) Phylogeography of ground tit (*Pseudopodoces humilis*) based on mtDNA: evidence of past fragmentation on the Tibetan Plateau. *Molecular Phylogenetics and Evolution*, **41**, 257–265.
- Young N, dePamphilis C (2005) Rate variation in parasitic plants: correlated and uncorrelated patterns among plastid genes of different function. *BMC Evolutionary Biology*, **5**, 16.
- 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.
- Zhang Q, Chiang T, George M, Liu J, Abbott R (2005) Phylogeography of the Qinghai-Tibetan Plateau endemic *Juniperus przewalskii* (Cupressaceae) inferred from chloroplast DNA sequence variation. *Molecular Ecology*, **14**, 3513–3524.
- Zhang YL, Li BY, Zheng D (2002) A discussion on the boundary and area of the Tibetan Plateau in China. *Geographical Research*, **21**, 1–10 (in Chinese with English abstract).
- Zheng D, Yao T (2004) *Uplifting of Tibetan Plateau with its Environmental Effects*. Science Press, Beijing, China.
- Zheng B, Xu Q, Shen Y (2002) The relationship between climate change and Quaternary glacial cycles on the Qinghai-Tibetan Plateau: review and speculation. *Quaternary International*, **97–98**, 93–101.

The authors collaborate on studies involved in molecular systematics and evolution, biogeography and speciation of seed plants. Xiao-Quan Wang is focusing on gymnosperm phylogeny and evolution, biogeography, speciation and floristic origin in the Qinghai-Tibetan Plateau. Fu-Sheng Yang is interested in adaptive radiation, population genetics and spatio-temporal distribution of various plants in the QTP. Yu-Fei Li and Xing Din are working on phylogeographic analysis on some alpine plants.
