*Botanical Journal of the Linnean Society*, 2008, **157**, 125–140. With 4 figures

# Potential refugium on the Qinghai–Tibet Plateau revealed by the chloroplast DNA phylogeography of the alpine species *Metagentiana striata* (Gentianaceae)

SHENGYUN CHEN[1,2], GUILI WU[3], DEJUN ZHANG[1,2], QINGBO GAO[1,2], YIZHONG DUAN[1,2], FAQI ZHANG[1,2] and SHILONG CHEN[1]*

[1]*Key Laboratory of Adaptation and Evolution of Plateau Biota, North-west Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810001, China*
[2]*Graduate School of the Chinese Academy of Sciences, Beijing 100039, China*
[3]*Key Laboratory of Arid and Grassland Ecology, Lanzhou University, Lanzhou 730000, China*

*Received 26 June 2007; accepted for publication 22 October 2007*

*Metagentiana striata* is an alpine annual herbaceous plant endemic to the east of the Qinghai–Tibet (Q–T) Plateau and adjacent areas. The phylogeography of *M. striata* was studied by sequencing the chloroplast DNA (cpDNA) *trn*S–*trn*G intergenic spacer. Ten haplotypes were identified from an investigation of 232 individuals of *M. striata* from 14 populations covering the entire geographical range of this species. The level of differentiation amongst populations was very high ($G_{ST} = 0.746$; $N_{ST} = 0.774$) and a significant phylogeographical structure was observed ($P < 0.05$). An analysis of molecular variance found a high variation amongst populations (76%), with $F_{ST} = 0.762$ (highly significant, $P < 0.001$), indicating that little gene flow occurred amongst the different regions; this was explained by the isolation of populations by high mountains along the Q–T Plateau and adjacent areas ($N_m = 0.156$). Only one ancestral haplotype (A) was common and widespread throughout the distributional range of *M. striata.* The populations of the Hengduan Mountains region of the south-eastern Q–T Plateau showed high diversity and uniqueness of haplotypes. It is suggested that this region was the potential refugium of *M. striata* during the Quaternary glaciation, and that interglacial and postglacial range expansion occurred from this refugium. This scenario was in good agreement with the results of nested clade analysis, which inferred that the current spatial distribution of cpDNA haplotypes and populations resulted from range expansion, together with past allopatric fragmentation events. © 2008 The Linnean Society of London, *Botanical Journal of the Linnean Society*, 2008, **157**, 125–140.

ADDITIONAL KEYWORDS: trnS-trnG – Hengduan Mountain – Qilianshan Mountains.

## INTRODUCTION

The present-day alpine flora of the Qinghai–Tibet (Q–T) Plateau comprises approximately 1816 species of seed plants in more than 339 genera. It is estimated that about 33 genera and 33.2% of the total species are endemic to this high and frigid region at 4200 m above sea-level (Wu, Yang & Fei, 1995; Wu & Wu, 1996). Little is known of the demographic history of alpine plants on the Q–T Plateau. The Q–T Plateau is the youngest, largest, and highest plateau on Earth, covering more than 2.5 million square kilometres at an average elevation of about 4000 m above sea-level (Zheng, 1996). It has been through many periods of uplift in its geological history. Its altitude reached as high as 3000 m in the Quaternary period, and it has maintained a trend of rapid uplift, with the most recent uplift event occurring since the Pliocene (Shi, Li & Li, 1998). The Q–T Plateau not only had an important influence on the atmospheric circulation of the northern hemisphere, but also directly affected the climatic and eco-environmental evolution of

*Corresponding author. E-mail: slchen@nwipb.ac.cn; chenshil@public.xn.qh.cn

China in the Quaternary period (Huairen & Xin, 1985; Zhang, Li & Ben, 2000). Data accumulated during investigations over the last four decades have indicated that three to four Quaternary glaciations occurred on the Q–T Plateau (Shi *et al*., 1995; Shi, Zheng & Yao, 1997; Shi, Huang & Yao, 2000; Zheng, Xu & Shen, 2002). In the Quaternary (approximately 2 million years ago, Ma), the distribution and composition of the Q–T flora was greatly affected not only by the cycles of glaciation and interglaciation, but also by the uplift of the Plateau. The Q–T flora showed an initial development in the Late Cretaceous or early Tertiary, and a period of major uplift of the Q–T Plateau occurred at about 3.6–1.7 Ma (Shi *et al*., 1998; Sun & Li, 2003). This latter interval provided the maximum time for the colonization of montane and alpine habitats from lowland areas and neighbouring regions. Although the Q–T Plateau is now mainly covered by alpine steppe and alpine desert, with various herbaceous taxa as the dominant components, its history is still unknown. During the Oligocene, the grasslands were already present, and gradually changed into semideserts; pollen records indicate that montane forests mixed with trees from warm- to cool-temperate regions were also present on the Q–T Plateau. The present alpine steppe and alpine desert developed during the late Pliocene with major uplift of the plateau (Wu, 1980; Zhang, 1983; Axelrod, Al-Shehbaz & Raven, 1996; Shi *et al*., 1998). However, there are other suggestions that the current dominant alpine steppe and alpine desert developed during the late Holocene, and that the alpine valley forest on the eastern rim of the Plateau expanded westward and upward during this period (Ren & Beug, 2002). Historical processes, such as bottlenecks, migrations, habitat fragmentation caused by vicariance, and genetic drift, are often reflected in the present-day genetic composition of populations, and permit the reconstruction of the historical biogeography of extant species (Hewitt, 1996, 2000, 2004; Stehlik *et al*., 2002).

The choice of tools for historical biogeographical investigations was limited before the development of molecular methods. Geographical mapping of macrofossils and pollen deposits documented large-scale plant migrations (Tang *et al*., 1998; Ren & Beug, 2002). However, the fossil record is limited in taxonomic resolution. In addition, fossils of herbaceous or alpine plants are especially scarce. Earlier studies on alpine plants mainly relied on their distribution pattern (Wu, 1980; Wu & Wu, 1996). Phylogeographical studies in plants using molecular markers have so far focused primarily on Europe (Hewitt, 2001; Abbott & Comes, 2004; Bartish, Kadereit & Comes, 2006). Considerable knowledge is now available for the flora of this continent, supporting detailed hypotheses on Pleistocene migration routes and possible ice age refugia Abbott *et al*., 2000; Hampe *et al*., 2003). Yet, there is little information available to address these questions for the Q–T Plateau plant species. Zhang *et al*. (2005) have suggested that the southeastern region of the Q–T Plateau was a possible refugium during the last glaciation from a phylogeographical study of a key tree species, *Juniperus przewalskii*, endemic to the Q–T Plateau region, and that range expansion was the major process influencing the present-day spatial distribution of haplotypes. However, phylogeographical studies on herbaceous plants of the Q–T Plateau are almost non-existent.

*Metagentiana striata* (Maxim.) T. N. Ho, S. W. Liu and S. L. Chen is an alpine annual herbaceous plant endemic to the east of the Q–T Plateau and adjacent areas (Ho & Pringle, 1995; Ho, Chen & Liu, 2002). Capsules are composed of many seeds without any special morphological adaptation for dispersal. The typical habitats of *M. striata* are alpine steppe and shrub slopes at altitudes of 2200–4500 m above sea-level (Ho & Pringle, 1995). *Metagentiana striata* is one of the most common and dominant herbaceous plant in the alpine zone in the east of the Q–T Plateau and adjacent areas. Thus, this species may be useful as a model for phylogeographical studies on alpine plants on the Q–T Plateau.

Since its development by Avise, Arnold & Ball (1987), phylogeography has become an increasingly important field of research within biogeography. The aim originally was to describe the distribution of genetic variation in space and time. More recently, the understanding of historical and population processes has emerged as a central focus (Abbott *et al*., 2000; Abbott & Comes, 2004; Zhang *et al*., 2005; Bartish *et al*., 2006). For this purpose, chloroplast DNA (cpDNA) markers are used, which, as a result of a lack of recombination and maternal inheritance in many angiosperms (Birky, Fuerst & Maruyama, 1989; Ennos, 1994; Martinez *et al*., 1997), are usually considered to provide a more conservative historical view than nuclear markers. In this paper, the intraspecific patterns of variation of the cpDNA non-coding fragment *trn*S (GCU)-*trn*G (UCC) intergenic spacer (abbreviated as *trn*S-*trn*G hereafter) (Hamilton, 1999) were investigated in all individuals of populations sampled throughout the geographical range of *M. striata*. The following questions were examined: (1) What is the phylogeographical structure of *M. striata*? (2) Does this structure allow for the inference of the glacial refugium and postglacial migration patterns of *M. striata* on the Q–T Plateau and adjacent areas?

## MATERIAL AND METHODS

### Population sampling

Fresh leaf material was sampled from the entire range of distribution of *M. striata* during the years 2003–04 (Table 1, Fig. 1), and was collected from about 5–21 individuals per population, with at least 5 m between individuals. In total, samples of 232 individuals from 14 populations were obtained. The collected leaf material was dried and stored in silica gel (Chase & Hills, 1991). Voucher specimens of all populations were deposited at the herbarium of the North-west Plateau Institute of Biology (HNWP), Xining, Qinghai Province, China.

### DNA extraction, polymerase chain reaction, and sequencing

Total genomic DNA was extracted from silica gel-dried leaf material using the 2 × cetyltrimethylammonium bromide (CTAB) procedure (Doyle & Doyle, 1987). Before all of the individuals were sequenced, 70 individuals sampled from 14 different populations (five individuals from each population) were initially investigated with four different pairs of universal primers scanning the cpDNA. The *psb*B-*psb*F, *rpl*20-5′*rps*12, *trn*S-*trn*G, and *trn*H-*psb*A intergenic regions were amplified and sequenced with these primers (Hamilton, 1999) in 25 μL reactions. The most sequence variations were found within the *trn*S-*trn*G region of the 70 individuals examined, and therefore this region was selected for the full analysis of the variation of *M. striata*.

Polymerase chain reactions (PCRs) contained 1 μL (*c.* 10–20 ng) of genomic DNA extract, 2.5 μL of 10 × PCR buffer (with 1.5 mM $MgCl_2$), 0.5 μL of 10 mM deoxynucleoside triphosphates (dNTPs), 1.25 μL of 5 pM of each primer, and 0.25 μL (1.25 U) of Taq DNA polymerase (CASarray, Shanghai, China) in a total volume of 25 μL in a Biometra thermal cycler (Tpersonal 48). The cycling profile was 5 min at 95 °C, followed by 35 cycles of 30 s at 95 °C, 30 s at 56 °C, 60 s at 72 °C, and a final extension of 72 °C for 6 min. PCR products were resolved electrophoretically on 1.5% agarose gels run at 200 V in 1 × TAE, visualized by staining with ethidium bromide, and photographed under ultraviolet light.

All successfully amplified DNA fragments were purified using a CASpure PCR Purification Kit, following the manufacturer's protocol (CASarray), prior to sequencing. The primers used for amplification were the same as those employed for sequencing. The sequencing reactions were programmed at 95 °C for 8 s, followed by 31 cycles of 95 °C for 15 s, 50 °C for 15 s, 60 °C for 1 min and 30 s, and a final extension of 60 °C for 1 min and 30 s, in a Biometra thermal cycler (Tpersonal 48) using a DYEnamic Dye Terminator Cycle Sequencing Kit (Amersham), but with the reaction volumes scaled down to 10 μL. The cycle sequencing products were cleaned using Autoseq 96 plates (Amersham), and analysed with a MegaBACE DNA Analysis System (Amersham Biosciences Corp.). All individuals sampled from each population and both strands of DNA were sequenced.

### Analysis of genetic indices

The cpDNA sequences of 232 individuals of 14 populations were aligned using the CLUSTAL X program (Thompson *et al.*, 1997), with additional minor manual adjustments. The different sequences (haplotypes) of cpDNA were identified using the DNASP 4.0 program (Rozas *et al.*, 2003). The indices of haplotypic diversity ($H_d$) and nucleotide diversity ($\pi$) (Nei, 1987) were calculated using the ARLEQUIN package (version 3.01; Excoffier, Laval & Schneider, 2006). Calculations of the average gene diversity within populations ($H_S$), total gene diversity ($H_T$), proportion of total diversity caused by differences between populations ($G_{ST}$), and the number of substitution types ($N_{ST}$) were performed according to Pons & Petit (1996) with the programs PERMUT 2 (http://www.pierroton.inra.fr/genetics/labo/software; 2000 permutations test) and HAPLODIV. $G_{ST}$ depends only on the haplotype frequency, whereas $N_{ST}$ is influenced by both the haplotype frequency and differences between haplotypes. A comparison was made between $N_{ST}$ and $G_{ST}$ using the U-statistic, which is approximated by a Gaussian variable by taking into account the covariance of the two values. If $N_{ST}$ is significantly higher than $G_{ST}$, closely related haplotypes occur in the same populations, indicating the presence of phylogeographical structure (Pons & Petit, 1996).

An analysis of molecular variance (AMOVA; Excoffier, Smouse & Quattro, 1992) was implemented using the ARLEQUIN package. The *F* statistic ($F_{ST}$) was calculated, and the significance was tested using 10 000 permutations; the average value of $N_m$ (gene flow) was estimated on the basis of the $F_{ST}$ value. A pairwise mismatch distribution was used to test for population expansion (Rogers & Harpending, 1992), and two neutrality tests with Tajima's *D* (Tajima, 1989) and Fu and Li's *D** (Fu & Li, 1993) were conducted using the DNASP 4.0 program. Tajima's D and Fu and Li's *D** indices were used to infer the nature of sequence evolution (for example, rapid selection or neutral) and probable historic population movements. Negative values are expected to occur when there has been recent population expansion (Fu, 1997; Knowles *et al.*, 1999) or a selective sweep (Fu, 1997; Crespi, Rissler & Browne, 2003). In contrast, positive values are expected when there has

**Table 1.** Population numbers, sample sizes, altitude, location, estimates of haplotype diversity ($H_d$), nucleotide diversity ($\pi$), and frequency of chloroplast DNA (cpDNA) haplotypes in the 14 populations of *Metagentiana striata* sampled in this study

| Population | Sample size | Location | Altitude (m) | No. of haplotype | $\pi$ | $H_d$ | Haplotype composition/frequency | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | A | B | C | D | E | F | G | H | I | J |
| 1. Pingan, Qinghai | 21 | 101°55′E, 36°20′N | 2690 | 1 | 0 | 0 | 21/1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 2. Maqing, Qinghai | 5 | 100°14′E, 34°47′N | 3520 | 1 | 0 | 0 | 5/1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 3. Banma, Qinghai | 20 | 100°47′E, 32°47′N | 3430 | 1 | 0 | 0 | 20/1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 4. Menyuan, Qinghai | 20 | 101°48′E, 37°17′N | 2940 | 1 | 0 | 0 | 20/1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 5. Menyuan, Qinghai | 12 | 102°00′E, 37°13′N | 2630 | 1 | 0 | 0 | 12/1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 6. Huzhu, Qinghai | 5 | 102°01′E, 37°13′N | 2640 | 1 | 0 | 0 | 5/1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 7. Aba, Sichuan | 15 | 101°18′E, 32°36′N | 3080 | 1 | 0 | 0 | 15/1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 8. Songpan, Sichuan | 20 | 103°34′E, 32°47′N | 2990 | 1 | 0 | 0 | 0.000 | 20/1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 9. Songpan, Sichuan | 20 | 103°31′E, 32°23′N | 2960 | 3 | 0.039 ± 0.020 | 0.679 ± 0.052 | 0.000 | 6/0.300 | 9/0.450 | 5/0.250 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 10. Hezuo, Gansu | 15 | 102°49′E, 34°54′N | 3210 | 2 | 0.002 ± 0.001 | 0.476 ± 0.092 | 5/0.333 | 0.000 | 0.000 | 0.000 | 10/0.667 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 11. Longde, Ningxia | 20 | 106°12′E, 35°40′N | 2740 | 1 | 0 | 0 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20/1.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 12. Hongyuan, Sichuan | 20 | 102°36′E, 32°24′N | 4300 | 3 | 0.005 ± 0.003 | 0.490 ± 0.116 | 14/0.700 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 3/0.150 | 3/0.150 | 0.000 | 0.000 |
| 13. HuangZhong, Qinghai | 20 | 101°38′E, 36°19′N | 3210 | 1 | 0 | 0 | 20/1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 14. Daofu, Sichuan | 19 | 101°16′E, 30°49′N | 3510 | 3 | 0.001 ± 0.001 | 0.292 ± 0.127 | 1/0.053 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 16/0.842 | 2/0.105 |
| Total | 232 | | | 10 | 0.014 ± 0.007 | 0.620 ± 0.034 | | | | | | | | | | |

© 2008 The Linnean Society of London, *Botanical Journal of the Linnean Society*, 2008, **157**, 125–140

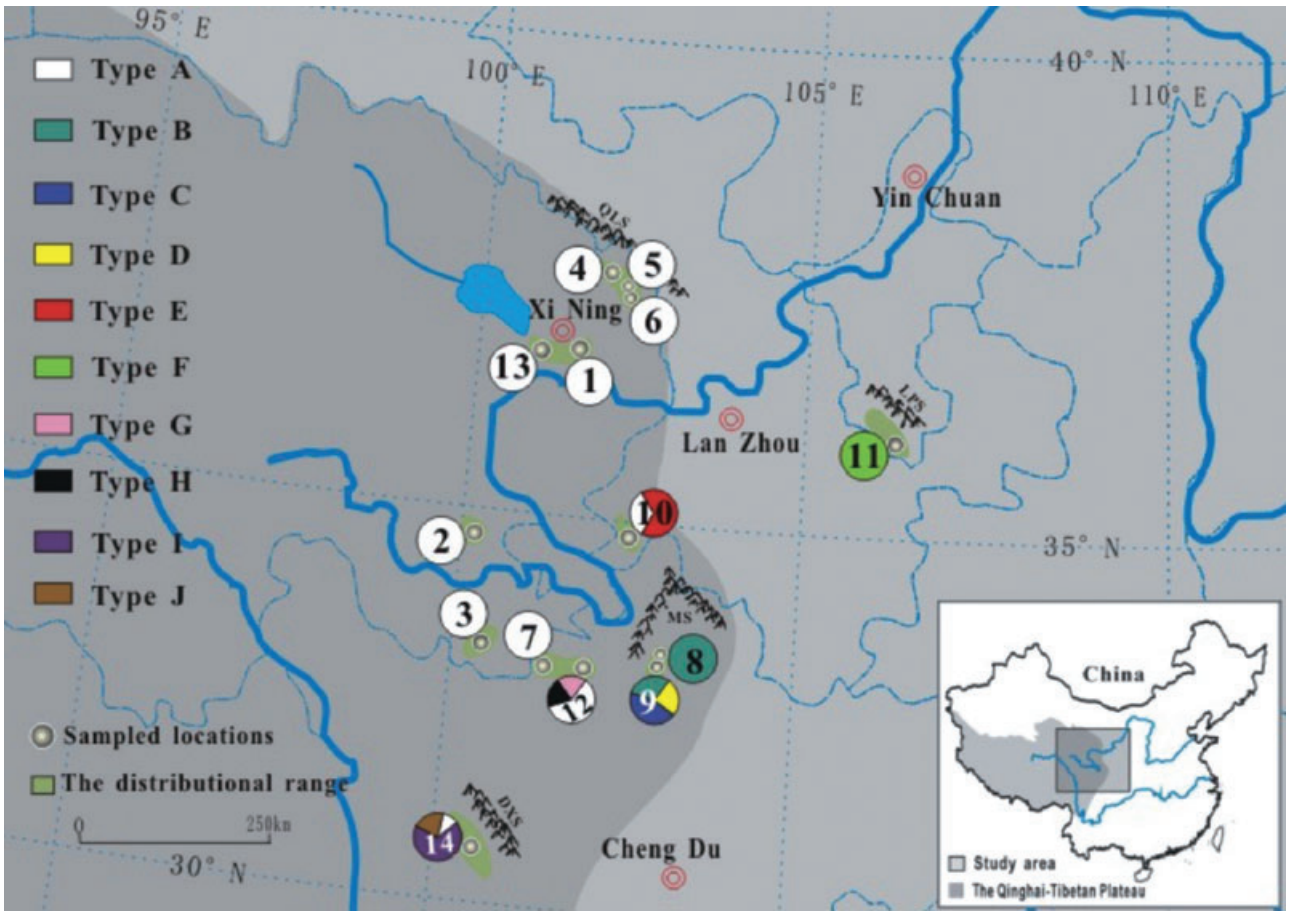


**Figure 1.** Sampling sites and geographical distribution of chloroplast DNA (cpDNA) haplotypes (A–J) detected in *Metagentiana striata*. The population numbers correspond to those detailed in Table 1. The darkish area indicates the Qinghai–Tibet Plateau region. The abbreviations QLS, MS, DXS, and LPS represent Qilianshan Mountains, Minshan Mountains, Daxueshan Mountains, and Liupanshan Mountains, respectively.

been population isolation in which long-term geographical subdivision enhances the accumulation of mutational differences between populations, or when balancing selection dominates (Rogers & Harpending, 1992; Marjoram & Donnelly, 1994).

### PHYLOGENETIC ANALYSIS

Phylogenetic relationships amongst the cpDNA haplotypes were evaluated by maximum parsimony (MP) and maximum likelihood (ML) analyses in PAUP* 4.0b10 (Swofford, 2003), and by Bayesian analysis with MrBayes 3.1.1 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003), using four species (*M. pterocalyx*, *M. rhodantha*, *Crawfurdia tibetica*, and *Tripterospermum cordatum*) as outgroups. Gaps were treated as missing in all analyses. In MP analysis, characters were equally weighted and unordered (Fitch, 1971). The program Modeltest, version 3.06 (Posada & Crandall, 1998) was used to select parameters and assumptions for ML analysis. Both MP and ML heuristic searches with 1000 random additions of sequence replicates, in combination with ACCTRAN character optimization, MULPARS, tree bisection–reconnection (TBR) branch swapping, and STEEPEST DESCENT on, were utilized to search for possible multiple islands of most parsimonious trees (Maddison, 1991). The relative support for relationships between haplotype clades was evaluated by bootstrap (BS) analysis (Felsenstein, 1985). BS values were calculated using 1000 replicates of heuristic searches, each with ten random addition sequence replicates using TBR and MULPARS on options. A Bayesian phylogeny was performed with GTR + $\Gamma$ + PINVAR parameters being estimated during the run, and using the default value of four Markov chains. Multiple chains can assist in the easier traversal of tree space and can help avoid entrapment in local topological optima. The Monte Carlo Markov

chain (MCMC) length was 1 000 000 generations, and the chain was sampled every 100 generations. Log-likelihood values for sampled trees stabilized after approximately 200 000 generations. Therefore, the last 8000 sampled trees were used to estimate Bayesian posterior probabilities (BPPs), also called Bayesian support values, after burn-in that contained each of the observed bipartitions (Larget & Simon, 1999).

### Haplotype network and nested clade analysis

Intraspecific relationships amongst the cpDNA haplotypes were constructed using the program NETWORK (Weir, 1996) with the aid of MINSPNET (Excoffier & Smouse, 1994). The haplotype network was nested following the procedure described in Templeton, Boerwinkle & Sing (1987) and Templeton, Crandall & Sing (1992) by first collapsing haplotypes at the tips of the networks (zero-step clade) one mutational step backward to form a one-step clade; this procedure was repeated until the entire nested cladogram was formed. For each clade (haplotype), the topological situation was specified with I or T, where I and T are the interior (ancestral) and tip (recent) clades or haplotypes. The resulting hierarchical structure of the clades reflects evolutionary time, with the lower nesting clades associated with more recent evolutionary events relative to the higher nesting clades (Crandall, 1996). To evaluate the relative contributions of historical events vs. population processes in shaping the observed genetic patterns, a nested clade analysis was conducted (Templeton, Routman & Phillips, 1995; Templeton, 1998) on the nested cladogram. This analysis examines the geographical association between haplotypes or clades and sampling locations, using both a simple categorical test and a more elaborate geographical distance test. The former test permutes clades against sample locations as categorical variables, excluding distance data. The latter test incorporates geographical distance data to estimate three statistics, namely within-clade distances ($D_c$) that measure the geographical spread of a clade, nested clade distances ($D_n$) that measure the geographical distance of a clade from the geographical centre of the nested clades, and contrast (I–T) between the interior and tip clades. The 95% significance level of these estimates was obtained through comparisons against those generated by a minimum of 10 000 random permutations. All calculations were performed in GEODIS 2.5 (Posada, Crandall & Templeton, 2000). The revised version of the inference key for nested clade analysis (Templeton, 2004) was used to infer whether the observed genetic clade patterns are best attributed to historical events, such as range expansion, or allopatric or past fragmentation.

## RESULTS

### Sequence variation

The lengths of the unaligned *trn*S-*trn*G sequences varied from 560 to 590 base pairs (bp) by screening 232 samples in 14 populations across the entire geographical range of *M. striata*. The total alignment of this region was 590 bp in length and identified ten different sequences (haplotypes). The ten cpDNA haplotypes were identified alphabetically (A–J), and the haplotype compositions and frequencies in each population are presented in Table 1 with geographical distributions illustrated in Figure 1. Variable sites amongst the ten haplotypes were observed with nine nucleotide substitutions and four indels (Table 2). The nine nucleotide substitutions were mainly four transitions of A↔G (157 bp site) and C↔T (sites at 442, 489, and 497 bp) and five transversions of A↔C (sites at 71 and 328 bp), T↔G (sites at 269 and 403 bp), and A↔T (366 bp site). Four indels of 10/13/1/20 bp were present from 89/315/328/381 bp to 98/327/328/400 bp, respectively. The nucleotide percentage of A and T was 69.5% for all sequences of *trn*S-*trn*G, and this phenomenon was in agreement with the nucleotide composition in most cpDNA intergenic spacers (Li, 1997; Chiang *et al*., 2001; Lu *et al*., 2002). The ten cpDNA haplotypic sequences, together with four outgroup individuals, are deposited in GenBank (Accession numbers EU552044-EU552053).

### Diversity indices and population differentiation

In the 14 populations of *M. striata*, haplotype A was widely distributed in eight populations and had the highest frequency (0.595), whereas haplotype J had the lowest frequency (0.009). The nucleotide diversity ($\pi$) within the 14 populations ranged from zero to 0.039, with a total value of 0.014 (Table 1). When compared with other plants that occur in Arctic-Alpine regions, the nucleotide diversity of *M. striata* ($\pi = 0.014$) was less than that of *Draba aizoides* ($\pi = 0.035$, from *trn*L-*trn*F analysis) (Widmer & Baltisberger, 1999) and *Vaccinium uliginosum* ($\pi = 0.063$, from *trn*L-*trn*F and *trn*S-*trn*G analyses) (Alsos *et al*., 2005). Population 9, located in the Hengduan Mountains region of the south-eastern Q–T Plateau, had the highest level of nucleotide diversity ($\pi = 0.039$), and population 12 in the Hengduan Mountains region also had high levels of nucleotide diversity ($\pi = 0.005$). The haplotypic diversities ($H_d$) within the 14 populations ranged from zero to 0.679, with a total value of 0.620 (Table 1). Population 9 containing three haplotypes was the most variable population ($H_d = 0.679$). Others, such as populations 10 ($H_d = 0.476$) and 12 ($H_d = 0.490$) in the Hengduan Mountains region, also contained high haplotypic diversities.

**Table 2.** Variable sites of the aligned sequences of the chloroplast DNA (cpDNA) fragment *trn*S-*trn*G in the ten haplotypes of *Metagentiana striata*. Sequences are numbered from the 5′ to the 3′ end in the region

| | trnS-trnG region Nucleotide position | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 7 1 | 8 9 | 1 5 7 | 2 6 9 | 3 1 5 | 3 2 8 | 3 6 6 | 3 8 1 | 4 0 3 | 4 4 2 | 4 8 9 | 4 9 7 |
| Haplotype | | | | | | | | | | | | |
| Type A | C | + | A | T | # | A | A | * | G | C | C | C |
| Type B | C | + | G | T | # | A | A | – | T | C | C | C |
| Type C | A | – | G | T | – | – | A | * | G | C | T | C |
| Type D | C | – | G | T | # | A | A | – | T | C | C | C |
| Type E | C | + | G | T | # | A | A | * | T | C | C | C |
| Type F | C | + | G | G | # | A | A | * | G | C | C | C |
| Type G | C | – | G | T | # | A | A | * | G | C | C | C |
| Type H | C | + | A | T | # | A | A | * | G | T | C | C |
| Type I | C | + | A | T | # | C | T | * | G | C | C | T |
| Type J | C | + | A | T | # | A | T | * | G | C | C | T |

The symbols +, #, *, and – denote different insertions/deletions: +, ATTATATAGA; #, AGATTCTTTTAAT; *, GGAATA CAAAATCTTCAAGC.

**Table 3.** Estimates of the average gene diversity within populations ($H_S$), total gene diversity ($H_T$), interpopulation differentiation ($G_{ST}$), and number of substitution types ($N_{ST}$) (standard errors in parentheses) within the total distribution of *Metagentiana striata* calculated with PERMUT2 (using a test with 2000 permutations) and HAPLODIV. $N_m$ (gene flow) was estimated by the ARLEQUIN program based on the $F_{ST}$ value (standard errors in parentheses)

| No. of populations | Arithmetic mean | Harmonic mean | No. of alleles | $H_S$ | $H_T$ | $G_{ST}$ | $N_{ST}$ | $N_m$ |
|---|---|---|---|---|---|---|---|---|
| 14 | 16.57 | 13.12 | 10 | 0.142 (0.064) | 0.558 (0.142) | 0.746 (0.089) | 0.774 (0.087) | 0.156 |

The average gene diversity within populations ($H_S$), total gene diversity ($H_T$), and the two coefficients of genetic differentiation ($G_{ST}$ and $N_{ST}$) over all populations were 0.142, 0.558, 0.746, and 0.774, respectively. A permutations test showed that $N_{ST}$ was significantly higher than $G_{ST}$ ($P < 0.05$) (Table 3). It indicated that the level of differentiation amongst populations was very high ($G_{ST} = 0.746$), that closely related haplotypes occurred in the same population, and the presence of a phylogeographical structure in the entire geographical range. AMOVA showed that about 76% of the genetic variation occurred amongst populations, whereas approximately 24% of the variation occurred within populations in the entire distribution region; the $F_{ST}$ value was 0.762 and highly significant ($P < 0.001$) (Table 4). At the species level, the estimated value of the average gene flow ($N_m$) amongst populations was very low ($N_m = 0.156$; see Table 3). The test values for neutrality on the total data set (Tajima's $D = -0.483$, $P > 0.10$; Fu and Li's $D^* = -0.385$, $P > 0.10$) were insignificant negative values, indicating that there may have been recent population expansion of *M. striata* on the Q–T Plateau and adjacent areas. The test of mismatch distribution indicated that it departed from the expectation by showing a single main peak (Fig. 2), which further confirmed that populations in the entire geographical region had undergone recent expansion.

### Phylogenetic analysis and nested clade analysis

The total aligned sequences of the ten haplotypes of cpDNA *trn*S-*trn*G, including the four outgroups, contained 676 sites, 592 of which were constant and 48 were parsimony informative. MP analysis produced two most parsimonious trees, with 89 steps, a consistency index (CI) of 0.966, and a retention index (RI) of 0.961. Figure 3 illustrates one most parsimonious tree, which has the same topology as the ML tree ($-\ln L = 1389.5823$) and the Bayesian majority rule consensus tree (not shown). Two clades with moderate

**Table 4.** Analysis of molecular variance (AMOVA) for 14 populations of *Metagentiana striata*

| Source of variation | d.f. | SS | VC | Variation (%) | Fixation index |
|---|---|---|---|---|---|
| Among populations | 13 | 54.530 | 0.250 | 76.18 | $F_{ST} = 0.762$* |
| Within populations | 218 | 17.065 | 0.078 | 23.82 | |
| Total | 231 | 71.595 | 0.329 | | |

d.f., degrees of freedom; SS, sum of squares; VC, variance components.
*$P < 0.001$.

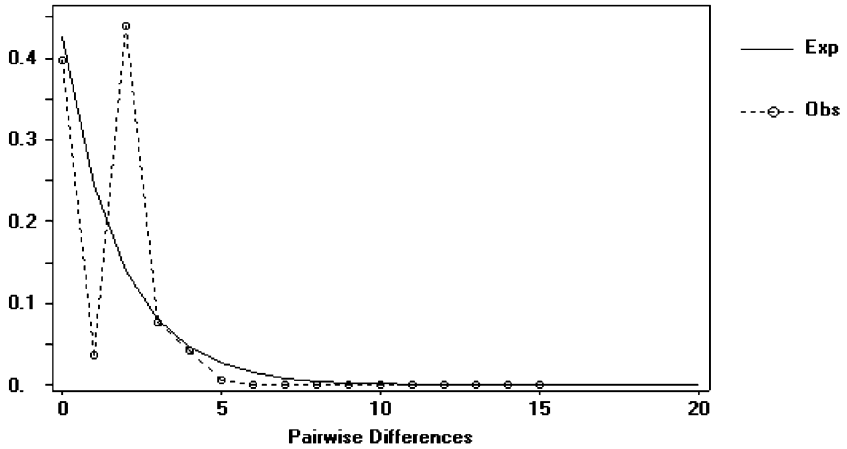


**Figure 2.** Mismatch distribution established for sequence data of the chloroplast DNA (cpDNA) fragment *trn*S-*trn*G from 232 individuals of *Metagentiana striata* in the entire sampling region. The thin line represents the expected mismatch distribution of a stationary population, and the broken line represents the observed mismatch distribution from segregation sites of the 232 aligned individual sequences of cpDNA *trn*S-*trn*G in *M. striata*.

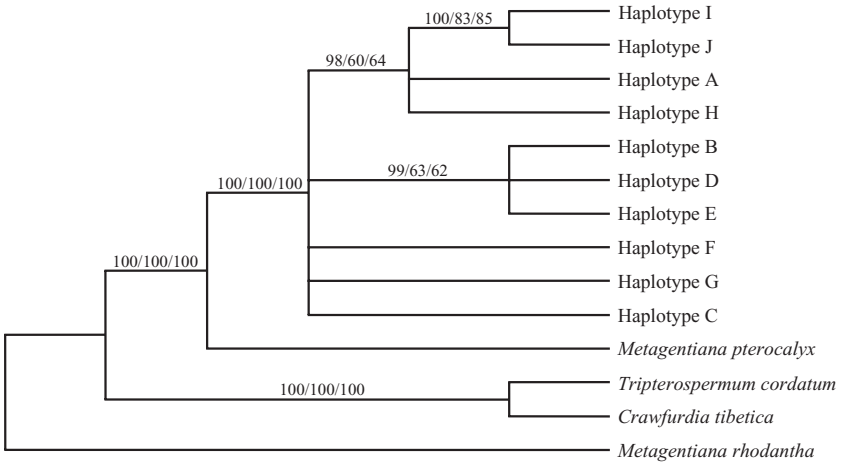


**Figure 3.** One of the two most parsimonious trees (length, 89; consistency index, 0.966; retention index, 0.961) based on the ten chloroplast DNA (cpDNA) haplotypes. This has the same topology as the 50% Bayesian majority rule consensus tree and best maximum likelihood tree (likelihood score, $-\ln L = 1389.5823$). Numbers above the branches indicate the Bayesian probabilities (BPPs), maximum parsimony and maximum likelihood bootstrap values, respectively, based on 1000 replicates.

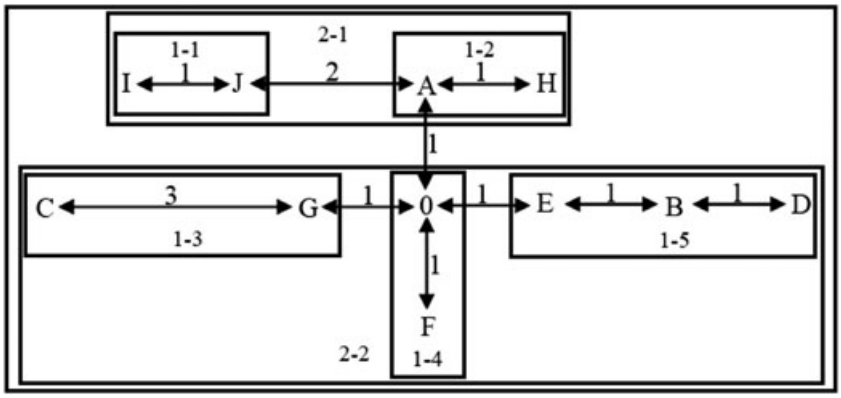


**Figure 4.** Nested cladogram of the chloroplast DNA (cpDNA) haplotypes (A–J) of *Metagentiana striata*. One-step clades are indicated by '1-*x*' and two-step clades by '2-*x*', where *x* is the number assigned to the clades within each level. '0' represents an absent haplotype, and the capital letters are the identified haplotypes. The numbers above or to the right of each connection represent the mutational steps.

BS support were detected in the most parsimonious tree (60% and 63%). However, relationships between the clades were not well resolved. One clade contained four haplotypes (A, H, I, and J) (98%, 60%, and 64%; Fig. 3). Haplotype A was dominant in populations of the north-eastern (Qilianshan Mountains, Fig. 1) and eastern regions of the Q–T Plateau, whereas haplotypes I and J were found only in population 14 of the Hengduan Mountains region on the south-eastern Q–T Plateau, and haplotype H was found only in population 12 of the Hengduan Mountains region (Fig. 1). Another clade contained haplotypes B, D, and E (99%, 63%, and 62%; Fig. 3). Haplotype E was dominant in population 10, whereas haplotype B and D were restricted to populations 8 and 9 in the Hengduan Mountains region (Fig. 1). The remaining haplotypes were not contained in any well-supported clades (Fig. 3). However, haplotype F was distributed only in population 11 of the adjacent alpine region (Liupanshan Mountains, Fig. 1) of the Q–T Plateau, whereas haplotypes C and G were found only in populations 9 and 12, respectively, of the Hengduan Mountains region (Fig. 1).

The nested cladogram of cpDNA haplotypes (A–J) of *M. striata* yielded two main clades (2-1 and 2-2; Fig. 4) that almost corresponded to the most parsimonious tree (Fig. 3). Clade 2-1 contains subclades 1-1 with haplotypes I and J, and subclade 1-2 with haplotypes A and H. The most common haplotype A is fixed in 11 populations of the entire geographical range. Clade 2-1 is connected to clade 2-2 by a one-step mutation. Clade 2-2 contains three subclades (1-3, 1-4, and 1-5). Haplotypes C and G are nested within subclade 1-3, haplotype F is nested within subclade 1-4, and haplotypes B, D, and E are nested within subclade 1-5. In this clade, the haplotypes, except for F, are distributed in the Hengduan Mountains region of the south-eastern Q–T Plateau. Obviously, the interior haplotypes A, E, and G are likely to represent ancestral haplotypes, and the other haplotypes (B, C, D, F, H, I, and J) were possibly derived recently in our present analysis. The results are presented in Table 5, including two estimations of geographical distance parameters ($D_c$, $D_n$) for each haplotype, clade, and contrast (I–T) at different cladistic levels, as well as significance values ($P$) in the nested clade analysis. According to the inference key, the corresponding chain of inferred historical events for each nested clade showing significant spatial geographical structure (as evidenced by two distance parameters; Table 5) is provided in Table 6. However, the results were ambiguous for clades 1-1 and 1-4 because statistically significant geographical structures were unresolved as a result of the rarity of the subclades (haplotypes). For the five significant clades, geographical sampling and tip/interior status were sufficient to yield conclusive outcomes. With regard to the entire nested clade, the historical event that shaped the present-day spatial distribution of all haplotypes was range expansion, whereas the distributions of haplotypes in clades 1-2 and 2-2 were shaped by past fragmentation, and in clades 1-3, 1-5, and 2-1 by allopatric fragmentation (Table 6).

## DISCUSSION

### PHYLOGEOGRAPHICAL STRUCTURE IN *M. STRIATA*

Although the cpDNA of Pinaceae and Cupressaceae was paternally inherited by pollen (Wagner, 1992; Zhang *et al.*, 2005; Meng *et al.*, 2007), other studies have suggested that cpDNA is maternally inherited via seed. The maternal inheritance leads to very low gene flow and high differentiation amongst

**Table 5.** Nested cladistic analysis of two geographical distances for chloroplast DNA (cpDNA) haplotypes (A–J) of *Metagentiana striata* based on the nested cladogram given in Figure 4

| Zero-step clades | | | | | | One-step clades | | | | | | Two-step clades | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Hap | Pos | $D_c$ | $P$ | $D_n$ | $P$ | Clade | Pos | $D_c$ | $P$ | $D_n$ | $P$ | Clade | Pos | $D_c$ | $P$ | $D_n$ | $P$ |
| I | T | | | | | 1-1 | T | $0.000_S$ | 0.000 | $444.45_L$ | 0.000 | 2-1 | T | $243.96_S$ | 0.003 | 257.26 | 0.298 |
| J | I | | | | | | | | | | | | | | | | |
| A | I | $211.08_S$ | 0.016 | $212.13_S$ | 0.009 | 1-2 | I | $213.10_S$ | 0.000 | $223.40_S$ | 0.000 | | | | | | |
| H | T | $0.00_S$ | 0.013 | $327.07_L$ | 0.008 | I-T | | $213.10_L$ | 0.000 | $-221.05_S$ | 0.000 | | | | | | |
| I–T | | $211.08_L$ | 0.013 | $-114.94_S$ | 0.008 | | | | | | | | | | | | |
| C | T | $0.00_S$ | 0.005 | $21.66_S$ | 0.005 | 1-3 | T | $32.48_S$ | 0.001 | 175.10 | 0.245 | 2-2 | I | $186.55_S$ | 0.000 | 249.60 | 0.257 |
| G | I | 0.00 | 0.387 | $64.96_L$ | 0.005 | | | | | | | I-T | | $-57.41_S$ | 0.000 | −7.66 | 0.269 |
| I–T | | 0.00 | 0.387 | $43.31_L$ | 0.005 | | | | | | | | | | | | |
| F | T | | | | | 1-4 | I | $0.00_S$ | 0.000 | $284.00_L$ | 0.000 | | | | | | |
| B | I | $16.18_S$ | 0.000 | $72.95_S$ | 0.000 | 1-5 | T | $109.13_S$ | 0.000 | $145.69_S$ | 0.000 | | | | | | |
| D | T | $0.00_S$ | 0.022 | 105.82 | 0.458 | I-T | | $-91.77_S$ | 0.004 | $131.65_L$ | 0.000 | | | | | | |
| E | I | $0.00_S$ | 0.000 | $180.92_L$ | 0.000 | | | | | | | | | | | | |
| I–T | | 11.68 | 0.394 | −2.88 | 0.402 | | | | | | | | | | | | |

Clades designated as in Figure 4. Hap, haplotypes; I, interior; I–T, average difference between interior and tip clades for both distance measures; Pos, position; T, tip.

Tests determine whether the within-clade ($D_c$) or nested clade ($D_n$) geographical distances are significantly large ($_L$) or significantly small ($_S$) based on 10 000 randomizations of the data, with the level of significance $P < 0.05$, where $P$ is the probability of a randomly generated value being equal to or larger (smaller) than the observed value.

**Table 6.** Chain of inference from the nested clade analysis of the chloroplast DNA (cpDNA) data set of *Metagentiana striata* following an updated version of Templeton's (2004) inference key for nested haplotype tree analysis of geographical distances. Permutational chi-squared probabilities for the geographical structure of the clades identified in Figure 4 from 10 000 resamples. *P* is the probability of a randomly generated chi-squared statistic being greater than or equal to the observed chi-squared

| Clade | Chi-squared statistic | *P* | Clade key | Inferences |
|---|---|---|---|---|
| 1-2 | 22.358 | 0.036 | 1-2-3-5-15-No | Past fragmentation |
| 1-3 | 12.000 | 0.005 | 1-19-No | Allopatric fragmentation |
| 1-5 | 55.336 | 0.000 | 1-19-No | Allopatric fragmentation |
| 2-1 | 149.563 | 0.000 | 1-19-No | Allopatric fragmentation |
| 2-2 | 107.074 | 0.000 | 1-2-3-5-15-No | Past fragmentation |
| Total cladogram | 204.718 | 0.000 | 1-2-11-12-No | Range expansion |

populations in many angiosperms (Ikeda *et al*., 2006; Koch *et al*., 2006). For the herbaceous species *M. striata*, very low average gene flow amongst populations ($N_m = 0.156$) and very high estimates of interpopulation differentiation ($G_{ST} = 0.746$; $N_{ST} = 0.774$) were recorded by examining the cpDNA haplotypes of 232 individuals from 14 populations (Table 3). This was consistent with the results of AMOVA (Table 4), which showed high genetic variation amongst populations (76%) and significant interpopulation differentiation ($F_{ST} = 0.762$; $P < 0.001$). Consequently, the results suggest that cpDNA shows a maternal mode of inheritance in *M. striata*. A very low gene flow occurred amongst populations because of the limited dispersal ability of seeds, as well as the vicariance and habitat fragmentation between populations by the high mountains of the Q–T Plateau and adjacent areas (Zhang *et al*., 1997). Moreover, $N_{ST}$ is significantly higher than $G_{ST}$ ($P < 0.05$), indicating the presence of a significant phylogeographical structure of *M. striata*, and that the closely related haplotypes occur in the same population in the entire geographical range. The related haplotypes A and H, B and D, and I and J co-occur in populations 12, 9, and 14, respectively, of the Hengduan Mountains region of the south-eastern Q–T Plateau.

### Demographic history of *M. striata*

In our study, the ancestral haplotypes (A, E, and G) existed together in the population of the Hengduan Mountains region of the south-eastern Q–T Plateau. The coexistence of several ancestral haplotypes in a restricted geographical area can result from two different, mutually non-exclusive scenarios. Firstly, it is possible that a particular region has accommodated populations over a long time, and that the ancestral haplotypes survived *in situ*. In this case, we would be dealing with a refugium. Secondly, it could be that the clades arrived postglacially by means of dispersal from one or several adjacent regions. In this case, we would be dealing with a recolonized area or an area of secondary contact (Pinceel *et al*., 2005). Although both scenarios might produce similar genetic signatures, it is argued that the first scenario (glacial refugium) is more parsimonious than the hypothesis of postglacial recolonization in the Hengduan Mountains region for the following reasons. Firstly, high cpDNA haplotype diversities were observed in most populations of the Hengduan Mountains region (population 9, $H_d = 0.679$; population 10, $H_d = 0.476$; population 12, $H_d = 0.490$; population 14, $H_d = 0.292$; Table 1), and there was very high haplotype uniqueness in this region (haplotypes E, C, D, G, H, I, and J; Fig. 1). Petit *et al*. (2003) concluded that plant populations in refugial areas show high genetic divergence and uniqueness. Secondly, the Hengduan Mountains range of the south-eastern Q–T Plateau is known not only as an important Tertiary centre of species diversification and a highly concentrated harbour of palaeo- and neo-endemics (Tao, 1992; Ying, Boufford & Zhang, 1993; Wang & Zhang, 1994), but also as an important glacial refugium of many plants (Wu, 1988; Nie *et al*., 2005; Zhang *et al*., 2005; Meng *et al*., 2007), including some taxa of Laurasian angiosperms, such as species of *Rhododendron*, *Rhodiola*, *Gentiana* (including *Metagentiana*; Ho *et al.*, 2002), and *Ciraeaster* (Wang, 1992). Therefore, it is suggested that the Hengduan Mountains region was possibly a potential refugium (supporting the first scenario) of *M. striata* during the Quaternary glaciation. In addition, most haplotypes of cpDNA were pooled in the Hengduan Mountains region, whereas only an ancestral haplotype (A) was distributed in the north-east (Qilianshan Mountains region) and east of the Q–T Plateau platform (Fig. 1). The populations (1, 2, 3, 4, 5, 6, 7, and 13) of *M. striata* with haplotype A might have experienced a series of bottlenecks following the founder effect to give a very low haplotype diversity ($H_d = 0$; Table 1) on the Q–T Plateau platform when recoloni-

zation occurred from the refugium (Hengduan Mountains region) during the interglaciation and postglaciation (Hewitt, 1996, 2000; Avise, 2000). Meanwhile, many recently derived haplotypes (for example, haplotypes B, D, H, and I) only showed one mutational step from ancestral haplotypes, indicating that these haplotypes possess limited dispersal ability and restricted regions (Hwang *et al*., 2003). The above scenario is supported by the suggestion that the Hengduan Mountains region of the south-eastern Q–T Plateau was the refugium of plants on the Plateau platform during the Quaternary glaciation, and these plants expanded from the refugium during interglacial and postglacial periods (Wu, 1979; Zhang *et al*., 2005).

One of the main origins of cpDNA evolution is from site mutations and short insertion/deletions (indels) (Clegg *et al*., 1994). In Table 2, there were nine nucleotide mutations and four short indels at 12 variable sites of the ten cpDNA haplotypes of the *trn*S-*trn*G fragment of *M. striata*. The results indicate that the intraspecies mutational rate is low, and that the genetic differentiation of *M. striata* may have occurred over a short time, as Chen *et al*. (2005) have suggested that the divergence time of *M. striata* was during the late Tertiary (about 6.2–3.3 Ma) based on the molecular clock hypothesis. Consequently, we propose that the ancestor species of *M. striata* developed in the Hengduan Mountains region during the late Tertiary (Chen *et al*., 2005) and occupied a wider distribution in the subsequent rapid radiation (northernmost to the Qilian Mountains and easternmost to the Liupan Mountains) (Fig. 1). This scenario is strongly supported by the present alpine steppe and alpine desert, which developed during the late Pliocene with major uplift of the Q–T Plateau (Wu, 1980; Zhang, 1983; Axelrod *et al*., 1996; Shi *et al*., 1998). However, the Quaternary glaciation might have resulted in the extinction of *M. striata* in the north-east (Qilian Mountains region) and east of the Q–T Plateau platform, but with survival in the refugium (Henduan Mountains region). During the interglacial and postglacial periods, range expansion took place from the refugium, and *M. striata* recolonized the Plateau platform. Analogous scenarios have been described by a series of studies (Hewitt, 1993; Lewis & Crawford, 1995; Templeton, 1998; Marshall, Newton & Ritland, 2002; Dobes, Mitchell-Olds & Koch, 2004; Cheng, Hwang & Lin, 2005; Watanabe, Kajita & Murata, 2006; Afzal-Rafii & Dodd, 2007).

The nested clade analysis also supported the above viewpoints. With regard to the entire nested cladogram, range expansion, inferred as a historical event, shaped the current spatial distribution of all haplotypes. Nested clade analysis indicated that, during the interglaciation and postglaciation, the populations of *M. striata* recolonized from the refugium to the Q–T Plateau platform (north-eastern and eastern Plateau) as a result of a range expansion event (Zhang *et al*., 2005). Likewise, the test results of Tajima's *D* and Fu and Li's *D** indices, which showed negative values, as well as the mismatch distribution of a single main peak, were completely in agreement with this inference, further confirming that recent population expansion has occurred in the entire geographical region. Past fragmentation events were also observed in clades 1-2 and 2-2 (Table 6). In clade 1-2, population 12 (with related haplotypes A and H) was isolated from population 10 (with haplotype A) of the north-eastern edge of the Hengduan Mountains region (Fig. 4) by large mountains (Minshan Mountains; Fig. 1). The large mountains blocked the gene flow via seeds between the two populations, as did the major uplift of the Q–T Plateau since the late Pliocene (Tao, 1992), and resulted in the past fragmentation event. With regard to clade 2-2, populations 8, 9, 10, and 12 of the Hengduan Mountains region and population 11 (unique haplotype F) of the Liupanshan Mountains region (Figs 1, 4) were separated by the large mountains of the Q–T Plateau and adjacent alpine region in the late Pliocene, resulting in the past fragmentation event. However, allopatric fragmentation events were inferred within clades 1-3, 1-5, and 2-1 (Table 6). From Figure 4, haplotypes C and G were nested in clade 1-3, haplotypes B, D, and E were nested in clade 1-5, and haplotypes A, H, I, and J were nested in two subclades (1-1 and 1-2) of clade 2-1. Based on the geographical distributions of these haplotypes (Fig. 1), it is suggested that allopatric fragmentation occurred amongst populations 9, 10, 12, and 14, isolated by the Minshan Mountains and Daxueshan Mountains (Fig. 1), during the interglaciation and postglaciation. In addition, population 8 with haplotype B was possibly derived from population 9 by range expansion. As stated above, range expansion from the Hengduan Mountains region (refugium) was the main historical process in the population of the Q–T Plateau platform, and allopatric and past fragmentation events have always occurred amongst populations in the Hengduan Mountains region and adjacent areas. The present spatial distributions of cpDNA haplotypes and populations of *M. striata* were shaped by our nested clade analysis.

Although the above findings on the herbaceous species *M. striata* were basically similar to the results obtained for two key tree species, *Juniperus przewalskii* and *Picea crassifolia*, which occur in the Q–T Plateau region (Zhang *et al*., 2005; Meng *et al*., 2007), this does not represent the demographic history of all species in the Q–T Plateau flora, because many earlier comparative studies have suggested that each species has its own unique history of glacial isolation

and interglacial or postglacial range expansion (Taberlet *et al*., 1998; Weider & Hobæk, 2000; Brochmann *et al*., 2003). Thus, in a strict sense, numerous species in the Q–T Plateau flora should be tested in future work.

## ACKNOWLEDGEMENTS

We gratefully acknowledge the comments and suggestions made by the reviewers. This research was supported by the National Science Foundation of China (30170066, 30370284).

## REFERENCES

**Abbott RJ, Comes HP. 2004.** Evolution in the Arctic: a phylogeographic analysis of the circumarctic plant, *Saxifraga oppositifolia* (purple saxifrage). *New Phytologist* **161:** 211–224.

**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.

**Afzal-Rafii Z, Dodd RS. 2007.** Chloroplast DNA supports a hypothesis of glacial refugia over postglacial recolonization in disjunct populations of black pine (*Pinus nigra*) in western Europe. *Molecular Ecology* **16:** 723–736.

**Alsos IG, Engelskjon T, Gielly L, Taberlet P, Brochmann C. 2005.** Impact of ice ages on circumpolar molecular diversity: insights from an ecological key species. *Molecular Ecology* **14:** 2739–2753.

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

**Avise JC, Arnold J, Ball RM. 1987.** Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* **18:** 489–522.

**Axelrod DI, Al-Shehbaz I, Raven PH. 1996.** History of the modern flora of China. In: Zhang A, Wu SG, eds. *Floristic characteristics and diversity of East Asian plants*. Beijing: China Higher Education Press, 43–55.

**Bartish IV, Kadereit JW, Comes HP. 2006.** Late Quaternary history of *Hippophae rhamnoides* L. (Elaeagnaceae) inferred from chalcone synthase intron (Chsi) sequences and chloroplast DNA variation. *Molecular Ecology* **15:** 4065–4083.

**Birky CM, Fuerst P, Maruyama T. 1989.** Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics* **121:** 613–627.

**Brochmann C, Gabrielsen TM, Nordal I, Landvik JY, Elven R. 2003.** Glacial survival or tabula rasa? The history of North Atlantic biota revisited. *Taxon* **52:** 417.

**Chase MW, Hills HH. 1991.** Silica gel: an ideal material for field preservation of leaf samples for DNA studies. *Taxon* **40:** 215–220.

**Chen S, Xia TAO, Wang Y, Liu J, Chen S. 2005.** Molecular systematics and biogeography of *Crawfurdia*, *Metagentiana* and *Tripterospermum* (Gentianaceae) based on nuclear ribosomal and plastid DNA sequences. *Annals of Botany* **96:** 413–424.

**Cheng Y-P, Hwang S-Y, Lin T-P. 2005.** Potential refugia in Taiwan revealed by the phylogeographical study of *Castanopsis carlesii* Hayata (Fagaceae). *Molecular Ecology* **14:** 2075–2085.

**Chiang TY, Chiang YC, Chen YJ, Chou CH, Havanond S, Hong TN, Huang S. 2001.** Phylogeography of *Kandelia candel* in East Asiatic mangroves based on nucleotide variation of chloroplast and mitochondrial DNAs. *Molecular Ecology* **10:** 2697.

**Clegg MT, Gaut BS, Learn GH, Morton RR. 1994.** Rates and pattern of chloroplast DNA evolution. *Proceedings of the National Academy of Sciences of the United States of America* **91:** 6795–6801.

**Crandall K. 1996.** Multiple interspecies transmissions of human and simian T-cell leukemia/lymphoma virus type I sequences. *Molecular Biology and Evolution* **13:** 115–131.

**Crespi EJ, Rissler LJ, Browne RA. 2003.** Testing Pleistocene refugia theory: phylogeographical analysis of *Desmognathus wrighti*, a high-elevation salamander in the southern Appalachians. *Molecular Ecology* **12:** 969–984.

**Dobes CH, Mitchell-Olds T, Koch MA. 2004.** Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American *Arabis drummondii*, *A.* × *divaricarpa*, and *A. holboellii* (Brassicaceae). *Molecular Ecology* **13:** 349–370.

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

**Ennos RA. 1994.** Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* **72:** 250–259.

**Excoffier L, Laval G, Schneider S. 2006.** *Arlequin 3.01: an integrated software package for population genetics data analysis*. Berne: Institute of Zoology, University of Berne.

**Excoffier L, Smouse PE. 1994.** Using allele frequencies and geographic subdivision to reconstruct gene trees with a species: molecular variance parsimony. *Genetics* **136:** 343–359.

**Excoffier LP, Smouse E, Quattro JM. 1992.** Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131:** 479–491.

**Felsenstein J. 1985.** Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39:** 783–791.

**Fitch WM. 1971.** Toward defining the course of evolution: minimum change for a specific tree topology. *Systematic Zoology* **20:** 406–416.

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

**Fu YX, Li WH. 1993.** Statistical tests of neutrality of mutations. *Genetics* **133:** 693–709.

**Hamilton MB. 1999.** Four primer pairs for the amplification

of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology* **8:** 521–523.

**Hampe A, Arroyo J, Jordano P, Petit RJ. 2003.** Range-wide phylogeography of a bird-dispersed Eurasian shrub: contrasting Mediterranean and temperate glacial refugia. *Molecular Ecology* **12:** 3415–3426.

**Hewitt GM. 1993.** Postglacial distribution and species substructure: lessons from pollen, insects and hybrid zones. In: Lees DR, Edwards D, eds. *Evolutionary patterns and processes*. London: Academic Press, 98–123.

**Hewitt GM. 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 GM. 2001.** Speciation, hybrid zones and phylogeography – or seeing genes in space and time. *Molecular Ecology* **10:** 537–549.

**Hewitt GM. 2004.** Biodiversity: a climate for colonization. *Heredity* **92:** 1–2.

**Ho TN, Chen SL, Liu SW. 2002.** *Metagentiana*, a new genus of Gentianaceae. *Botanical Bulletin of Academia Sinica* **43:** 83–91.

**Ho TN, Pringle JS. 1995.** Gentianaceae. In: Wu ZY, Raven PH, eds. *Flora of China*. Beijing/St. Louis, MO: Science Press/Missouri Botanical Garden, 1–140.

**Huairen Y, Xin X. 1985.** The influence of Quaternary climate on Chinese loess and desert. In: Li T, ed. *Contribution to the Quaternary glaciology and Quaternary geology*. Beijing: Geological Publishing House, 126–134.

**Huelsenbeck JP, Ronquist F. 2001.** MrBayes: Bayesian inference of phylogeny. *Bioinformatics* **17:** 754–755.

**Hwang S-Y, Lin T-P, Ma C-S, Lin C-L, Jeng-Der C, Yang J-C. 2003.** Postglacial population growth of *Cunninghamia konishii* (Cupressaceae) inferred from phylogeographical and mismatch analysis of chloroplast DNA variation. *Molecular Ecology* **12:** 2689–2695.

**Ikeda H, Senni KEI, Fujii N, Setoguchi H. 2006.** Refugia of *Potentilla matsumurae* (Rosaceae) located at high mountains in the Japanese archipelago. *Molecular Ecology* **15:** 3731–3740.

**Knowles LL, Futuyma DJ, Eanes WF, Rannala B. 1999.** Insight into speciation from historical demography in the phytophagous beetle genus *Ophraella*. *Evolution* **53:** 1846–1856.

**Koch MA, Kiefer C, Ehrich D, Vogel J, Brochmann C, Mummenhoff K. 2006.** Three times out of Asia Minor: the phylogeography of *Arabis alpina* L. (Brassicaceae). *Molecular Ecology* **15:** 825–839.

**Larget B, Simon DL. 1999.** Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Molecular Biology and Evolution* **16:** 750–759.

**Lewis PO, Crawford DJ. 1995.** Pleistocene refugium endemics exhibit greater allozymic diversity than widespread congeners in the genus *Polygonella* (Polygonaceae). *American Journal of Botany* **82:** 141–149.

**Li WH. 1997.** *Molecular evolution*. Sunderland, MA: Sinauer.

**Lu SY, Hong KH, Liu SL, Cheng YP, Wu WL, Chiang TY. 2002.** Genetic variation and population differentiation of *Michelia formosana* (Magnoliaceae) based on cpDNA variation and RAPD fingerprints: relevance to post-Pleistocene recolonization. *Journal of Plant Research* **115:** 203–216.

**Maddison DR. 1991.** The discovery and importance of multiple islands of most-parsimonious trees. *Systematic Zoology* **40:** 315–318.

**Marjoram P, Donnelly P. 1994.** Pairwise comparisons of mitochondrial DNA sequences in subdivided populations and implications for early human evolution. *Genetics* **136:** 673–683.

**Marshall HD, Newton C, Ritland K. 2002.** Chloroplast phylogeography and evolution of highly polymorphic microsatellites in lodgepole pine (*Pinus contorta*). *Theoretical and Applied Genetics* **104:** 367–378.

**Martinez P, Lopez C, Roldan M, Sabater B, Martin M. 1997.** Plastid DNA of five ecotypes of *Arabidopsis thaliana*: sequence of ndhG gene and maternal inheritance. *Plant Science* **123:** 113–122.

**Meng LH, Yang R, Abbott RJ, Miehe G, Hu TH, Liu JQ. 2007.** Mitochondrial and chloroplast phylogeography of *Picea crassifolia* Kom. (Pinaceae) in the Qinghai–Tibetan Plateau and adjacent highlands. *Molecular Ecology* **16:** 412.

**Nei M. 1987.** *Molecular evolutionary genetics*. New York: Columbia University Press.

**Nie ZL, Wen J, Gu ZJ, Boufford DE, Sun H. 2005.** Polyploidy in the flora of the Hengduan mountains hotspot, southwestern China. *Annals of the Missouri Botanical Garden* **92:** 275–306.

**Petit RJ, Aguiagalde I, de Beaulieu JL, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Aparajita M, Müller-Starck G, Demesure-Musch B, Palmé A, Martín JP, Rendell S, Vendramin GG. 2003.** Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300:** 1563–1565.

**Pinceel JAN, Jordaens K, Pfenninger M, Backeljau T. 2005.** Rangewide phylogeography of a terrestrial slug in Europe: evidence for Alpine refugia and rapid colonization after the Pleistocene glaciations. *Molecular Ecology* **14:** 1133–1150.

**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.

**Posada D, Crandall KA, Templeton AR. 2000.** GEODIS: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology* **9:** 487–488.

**Ren GY, Beug HJ. 2002.** Mapping Holocene pollen data and vegetation of China. *Quaternary Science Reviews* **21:** 1395–1422.

**Rogers AR, Harpending H. 1992.** Population growth makes waves in the distribution of pairwise differences. *Molecular Biology and Evolution* **9:** 552–569.

**Ronquist F, Huelsenbeck JP. 2003.** MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19:** 1572–1574.

**Rozas J, De Sanchez I, Barrie JC, Messeguer X, Rozas R. 2003.** DNASP: DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19:** 2496–2497.

**Shi YF, Huang M, Yao T. 2000.** *Glaciers and their environments in China*. Beijing: Science Press.

**Shi YF, Li JJ, Li BY. 1998.** *Uplift and environmental changes of Qinghai–Tibetan Plateau in the late Cenozoic*. Guangzhou: Guangdong Science and Technology Press.

**Shi YF, Zheng B, Li S, Ye B. 1995.** Studies on altitude and climatic environment in the middle and east parts of Tibetan Plateau during Quaternary maximum glaciation. *Journal of Glaciology and Geocryology* **17:** 97–112.

**Shi YF, Zheng B, Yao T. 1997.** Glaciers and environments during the Last Glacial Maximum (LGM) on the Tibetan Plateau. *Journal of Glaciology and Geocryology* **19:** 97–113.

**Stehlik I, Blattner FR, Holderegger R, Bachmann K. 2002.** Nunatak survival of the high Alpine plant *Eritrichium nanum* (L.) Gaudin in the central Alps during the ice ages. *Molecular Ecology* **11:** 2027–2036.

**Sun H, Li ZM. 2003.** Qinghai–Tibet plateau uplift and its impact on Tethys flora. *Advance in Earth Sciences* **18:** 852–862.

**Swofford DL. 2003.** *PAUP*: phylogenetic analysis using parsimony (*and other methods)*, version 4.0b10. Sunderland, MA: Sinauer Associates.

**Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F. 1998.** Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7:** 453–464.

**Tajima F. 1989.** Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **7:** 381–397.

**Tang L, Shen C, Kong Z, Wang F, Liu K. 1998.** Pollen evidence of climate during the last glacial maximum in eastern Tibetan Plateau. *Journal of Glaciology and Geocryology* **20:** 133–140.

**Tao JR. 1992.** The Tertiary vegetation and flora and floristic regions in China. *Acta Phytotaxonomica Sinica* **30:** 25–43.

**Templeton AR. 1998.** Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology* **7:** 381–397.

**Templeton AR. 2004.** Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology* **13:** 789–809.

**Templeton AR, Boerwinkle E, Sing CF. 1987.** A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics* **117:** 343–351.

**Templeton AR, Crandall KA, Sing CF. 1992.** A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132:** 619–633.

**Templeton AR, Routman E, Phillips CA. 1995.** Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the Tiger Salamander, *Ambystoma tigrinurn*. *Genetics* **140:** 767–778.

**Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997.** The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25:** 4876–4882.

**Wagner DB. 1992.** Nuclear, chloroplast and mitochondrial DNA polymorphisms as biochemical markers in population genetic analysis of forest trees. *New Forests* **6:** 373–390.

**Wang HS, Zhang YL. 1994.** The biodiversity and characters of spermatophytic genera endemic to China. *Acta Botanica Yunnanica* **16:** 209–220.

**Wang WT. 1992.** On some distribution patterns and some migration routes found in the Eastern Asiatic region (cont'd). *Acta Phytotaxonomica Sinica* **30:** 97–117.

**Watanabe K, Kajita T, Murata J. 2006.** Chloroplast DNA variation and geographical structure of the *Aristolochia kaempferi* group (Aristolochiaceae). *American Journal of Botany* **93:** 442–453.

**Weider LJ, Hobæk A. 2000.** Phylogeography and arctic biodiversity: a review. *Annales Zoologici Fennici* **37:** 217–231.

**Weir BS. 1996.** *Genetic data analysis II*. Sunderland, MA: Sinauer Associates.

**Widmer A, Baltisberger M. 1999.** Extensive intraspecific chloroplast DNA (cpDNA) variation in the alpine *Draba aizoides* L. (Brassicaceae): haplotype relationships and population structure. *Molecular Ecology* **8:** 1405–1415.

**Wu CY. 1979.** The regionalization of Chinese flora. *Acta Botanica Yunnanica* **1:** 1–22.

**Wu CY. 1980.** *The vegetation of China*. Beijing: Science Press.

**Wu CY. 1988.** Hengduan mountain flora and her significance. *Journal of Japanese Botany* **63:** 297–311.

**Wu CY, Wu SG. 1996.** A proposal for a new floristic kingdom (realm) – the E. Asiatic kingdom, its delineation and characteristics. In: Zhang A, Wu SG, eds. *Floristic characteristics and diversity of East Asian plants*. Beijing: China Higher Education Press, 3–42.

**Wu SG, Yang YP, Fei Y. 1995.** On the flora of the alpine region in the Qinghai–Xizang (Tibet) Plateau. *Acta Botanica Yunnanica* **17:** 233–250.

**Ying TS, Boufford DE, Zhang YL. 1993.** *The endemic genera of seed plants of China*. Beijing: Science Press.

**Zhang DF, Li FQ, Ben JM. 2000.** Eco-environmental effects of the Qinghai–Tibet Plateau uplift during the Quaternary in China. *Environmental Geology* **39:** 1352–1358.

**Zhang Q, Chiang TY, George M, Liu JQ, Abbott RJ. 2005.** Phylogeography of the Qinghai–Tibetan Plateau endemic *Juniperus przewalskii* (Cupressaceae) inferred from chloroplast DNA sequence variation. *Molecular Ecology* **14:** 3513–3524.

**Zhang RZ, Jin SK, Quan GQ, Li SH, Ye ZY, Wang PG, Zhang ML. 1997.** *Distribution of mammalian species in China*. Beijing: China Forestry Publishing House.

**Zhang XS. 1983.** The Tibetan Plateau in relation to the vegetation of China. *Annals of the Missouri Botanical Garden* **70:** 564–570.

**Zheng BX, Xu QQ, Shen YP. 2002.** The relationship between climate change and Quaternary glacial cycles on the Qinghai–Tibetan Plateau: review and speculation. *Quaternary International* **97:** 93–101.

**Zheng D. 1996.** The system of physico-geographical regions of the Qinghai–Tibet (Xizang) Plateau. *Science in China (Series D)* **39:** 410–417.