

RESEARCH ARTICLE

Open Access

# Phylogeography and allopatric divergence of cypress species (*Cupressus* L.) in the Qinghai-Tibetan Plateau and adjacent regions

Tingting Xu<sup>1</sup>, Richard J Abbott<sup>2</sup>, Richard I Milne<sup>3</sup>, Kangshan Mao<sup>1</sup>, Fang K Du<sup>1</sup>, Guili Wu<sup>1</sup>, Zhaxi Ciren<sup>4</sup>, Georg Miehle<sup>5</sup> and Jianquan Liu\*<sup>1</sup>

## Abstract

**Background:** Although allopatric speciation is viewed as the most common way in which species originate, allopatric divergence among a group of closely related species has rarely been examined at the population level through phylogeographic analysis. Here we report such a case study on eight putative cypress (*Cupressus*) species, which each have a mainly allopatric distribution in the Qinghai-Tibetan Plateau (QTP) and adjacent regions. The analysis involved sequencing three plastid DNA fragments (*trnD-trnT*, *trnS-trnG* and *trnL-trnF*) in 371 individuals sampled from populations at 66 localities.

**Results:** Both phylogenetic and network analyses showed that most DNA haplotypes recovered or haplotype-clustered lineages resolved were largely species-specific. Across all species, significant phylogeographic structure ( $N_{ST} > G_{ST}$ ,  $P < 0.05$ ) implied a high correlation between haplotypes/lineages and geographic distribution. Two species, *C. duclouxiana* and *C. chengiana*, which are distributed in the eastern QTP region, contained more haplotypes and higher diversity than five species with restricted distributions in the western highlands of the QTP. The remaining species, *C. funebris*, is widely cultivated and contained very little cpDNA diversity.

**Conclusions:** It is concluded that the formation of high mountain barriers separating deep valleys in the QTP and adjacent regions caused by various uplifts of the plateau since the early Miocene most likely promoted allopatric divergence in *Cupressus* by restricting gene flow and fixing local, species-specific haplotypes in geographically isolated populations. The low levels of intraspecific diversity present in most species might stem from population bottlenecks brought about by recurrent periods of unfavorable climate and more recently by the negative impacts of human activities on species' distributions. Our findings shed new light on the importance of geographical isolation caused by the uplift of the QTP on the development of high plant species diversity in the QTP biodiversity hotspot.

## Background

Geographic isolation between populations within species is a major precursor of allopatric speciation [1,2]. In the absence of gene flow, different alleles become fixed in different populations through the action of selection and/or genetic drift potentially leading to the origin of new species [3]. It is expected, therefore, that in regions where geographic isolation is common, for example, in areas where high mountains separate populations occurring in

deep valleys, allopatric divergence will be promoted and accelerated due to the formation of locally adapted gene pools and possibly the origin of intrinsic postzygotic isolation according to the Bateson-Dobzhansky-Muller model [3-5]. A phylogeographic analysis of closely related species that occur in such areas can be instructive on the importance of geographic isolation and allopatric divergence in their origin and subsequent evolution [6].

The Qinghai-Tibetan Plateau (QTP) and adjacent regions comprise an excellent mountainous system for studying plant diversification and evolution. Levels of plant diversity and endemism in these regions are very high and some core areas are regarded as biodiversity

\* Correspondence: ljdx@public.xn.qh.cn

<sup>1</sup> Division of Molecular Ecology, Key Laboratory of Arid and Grassland Ecology, School of Life Science, Lanzhou University, Lanzhou, 730000, China  
Full list of author information is available at the end of the article

hotspots (e.g. [7]). A few studies in this region have begun to elucidate the process of species diversification and/or trace shifts in the ranges of single species during the Quaternary climatic oscillations (e.g. [8,9]). However, no previous studies have attempted to examine genetic differentiation in this region among species of a closely related group, and in so doing shed light on their respective evolutionary histories. Here we report such a case study of a group of cypress species (*Cupressus* L.) that occurs on the QTP and adjacent areas of China and the Himalayas. Cupresses are trees or shrubs native to scattered localities in temperate regions in the northern hemisphere. Because of their attractive morphology they are important in horticulture and are widely cultivated in many parts of the world [10]. In several mountainous regions of the QTP, cupresses form natural forests [11]. From the results of a recent molecular phylogenetic analysis of cupresses, Little [12] concluded that True Cupresses only encompass species in the Old World, and that those in the New World should be transferred to *Callitropsis*. As a result, the genus *Cupressus* has been re-circumscribed to comprise 12 Old World species, distributed respectively in the Mediterranean region and Africa (*C. sempervirens*, *C. dupreziana* and *C. atlantica*) and Asia (*C. torulosa*, *C. gigantea*, *C. cashmeriana* and *C. austrotibetica* in the high altitude region of the QTP and west Himalayas, and *C. duclouxiana*, *C. chengiana*, *C. funebris*, *C. jiangensis* and *C. tonkinensis* in the low altitude region of the eastern QTP, central China and Vietnam [12]). Morphological distinction between some of these species is marginal causing taxonomic delimitation to be disputed, especially between Asian species [10,12-19]. We followed Little [12] and treated all taxonomic units recognized by him as separate species. Most of these Asian species occur allopatrically in the QTP and adjacent regions, and only a few are parapatric in parts of their range (Fig. 1) (Additional File 1) [10,15].

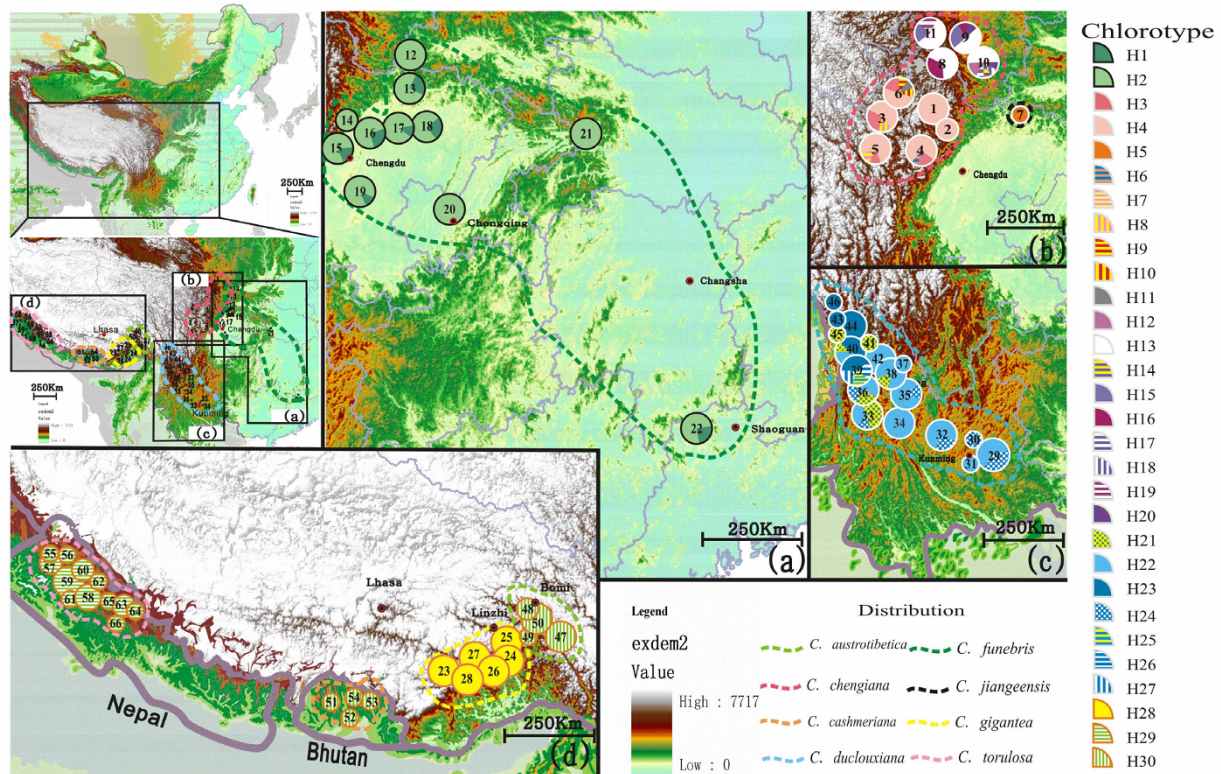
We used plastid (mainly chloroplast) DNA sequence data to examine interspecific divergence and to assess intraspecific diversity of these species. Plastid DNA is paternally inherited and dispersed mainly by pollen in all species of the Cupressaceae investigated [20-23]. Recent studies suggest that paternally inherited plastid DNA sequences in conifers may be species-specific and therefore effective in tracing lineage sorting between species [24-27]. Plastid DNA sequence variation has been used previously in phylogenetic studies of cupresses, but only single sample from most species were examined [12]. In addition, surveys of RAPD variation [28] and nuclear ITS variation [29] have been conducted within and among cypress species including those examined in the current study. However, the results of these studies were inconsistent with regard to interspecific relationships [12]. In this study, we aimed to address the following questions based

on surveys of sequence variation for three plastid DNA fragments within and between populations at the species level: (1) Do plastid DNA haplotypes or clusters of such haplotypes correspond with morphological differentiation and the allopatric distributions of the *Cupressus* species examined? (2) Does intraspecific DNA diversity correlate with the natural population size of a species and, in turn, provide guidance for conservation management?

## Results

### Haplotype distribution and relationships

In total, 30 haplotypes were identified (Additional Files 1, 2 and 3, Fig. 1), of which 17 occurred in *C. chengiana* (N = 89), seven in *C. duclouxiana* (N = 86), two in *C. funebris* (N = 89) and four among *C. gigantea* (N = 58), *C. jiangensis* (N = 1), *C. austrotibetica* (N = 22), *C. cashmeriana* (N = 4) and *C. torulosa* (N = 22). No intraspecific cpDNA variation was recorded in *C. gigantea*, *C. jiangensis*, *C. cashmeriana*, *C. austrotibetica* and *C. torulosa*. *Cupressus cashmeriana* and *C. austrotibetica* shared the same haplotype, which differed by a single nucleotide from the haplotype possessed by *C. torulosa*. These two sequences were very different from the haplotype possessed by *C. gigantea*. A neighbour joining analyses showed that the 30 haplotypes clustered into six different lineages (A-F) with most haplotypes being species-specific (Fig. 2). Although all six lineages were recovered in the 50% strict consensus tree of 51267 most parsimonious trees (CI = 0.77, RI = 0.87, Length = 141), the pattern of relationships between lineages received low bootstrap support (Fig. 2). Lineage A comprised nine haplotypes possessed by individuals of *C. chengiana* sampled from Gansu province, China. In contrast, the other eight haplotypes detected in this species were present only in individuals sampled from Sichuan, China, which together with the single haplotype (H5) recovered from *C. jiangensis*, comprised lineage E. The sister relationship between lineages A and E was poorly supported. Lineage D was composed of the single haplotype present in *C. torulosa* plus the haplotype shared by *C. austrotibetica* and *C. cashmeriana*, and was sister to lineages A and E (Fig. 2) of *C. chengiana*. The seven different haplotypes detected in *C. duclouxiana* formed a monophyletic lineage C, which was sister to lineage B that consisted of the single haplotype possessed by *C. gigantea*. Lineage F comprised the two haplotypes specific to *C. funebris* and was sister to the five other lineages (A to E) and to two cypress species occurring out of Asia. To evaluate the origins of these lineages (species), we selected the commonest haplotype representing each lineage and dated their origins. According to the estimated nucleotide mutation rate for *Cupressus* (Additional File 4), *C. funebris* was estimated to have diverged from the other lineage/species



**Figure 1** The distribution of eight cypress species and frequencies of plastid DNA haplotypes in each population surveyed. Upper left panel, a, b, c and d: Geographic ranges of eight cypress species and locations of populations sampled. The upper right panels a, b, c and lower panel d: plastid DNA haplotypes identified and the proportion of each present within populations of each species examined (circle size is proportional to the number of individuals surveyed per population). Further details of size and haplotype diversity of populations are provided in Additional File 1.

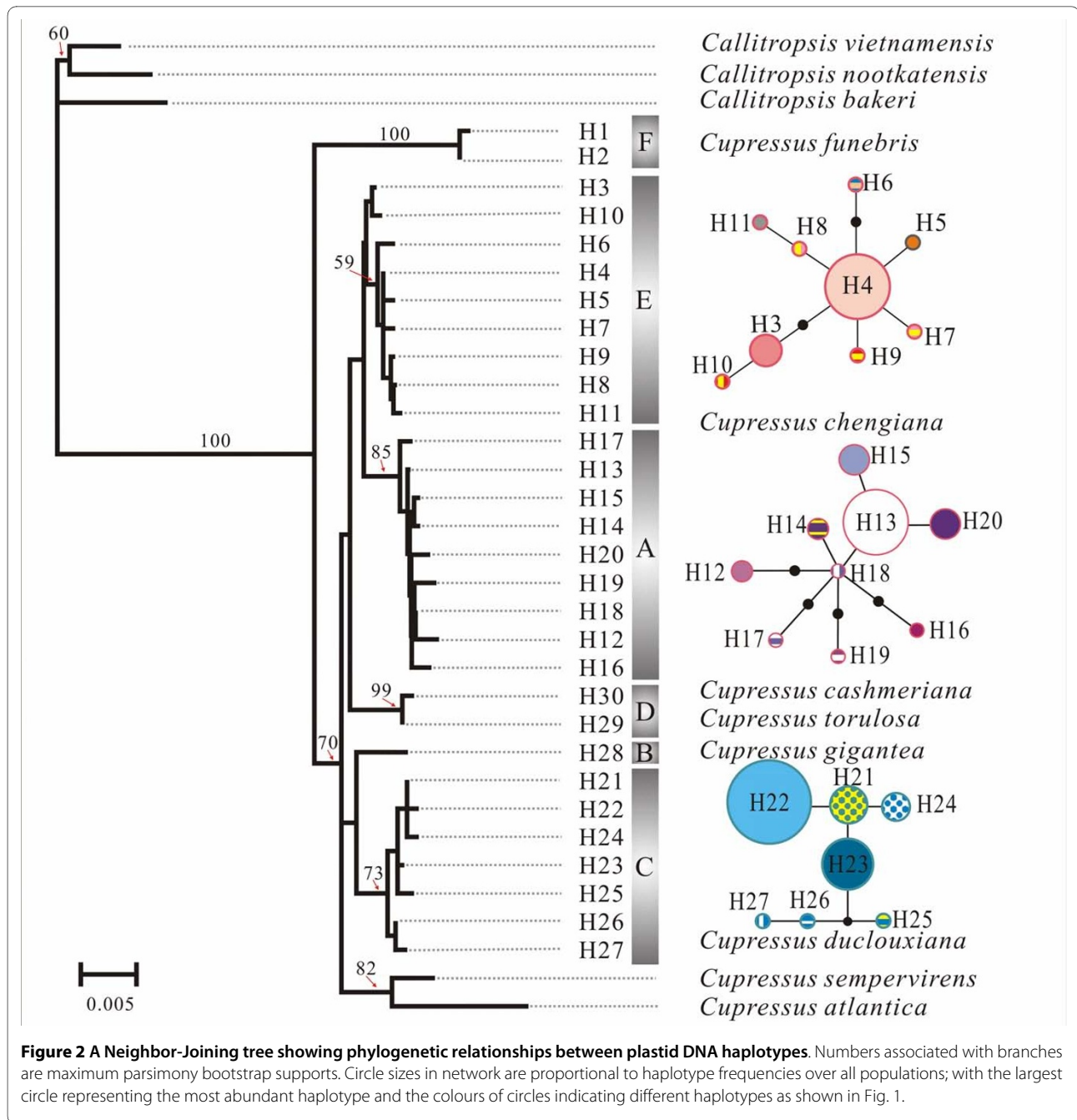
approximately 26 million years ago (Mya) (Additional File 4). *C. duclouxiana* in the eastern QTP diverged from the western *C. gigantea* around 10 Mya, while the eastern *C. chengiana* diverged from the western *C. torulosa/C. austrotibetica/C. cashmeriana* lineage around 13 Mya. The intraspecific divergence that gave rise to the two different lineages of *C. chengiana* identified was dated to approximately nine Mya.

Because the number of mutations distinguishing any two lineages was always greater than six, we constructed separate haplotype networks for the three lineages that contained most haplotypes (i.e. lineages A, C and E). Within lineage E, two haplotypes H10 and H3 occurred in a small sub-lineage that was separated by four mutations from other haplotypes (Fig. 2). However, H10 co-occurred with the most frequent haplotype in this lineage, H4, in most sampled populations. The single haplotype (H5) recovered from *C. jiangeensis* was also present in lineage E and is derived from H4. In lineage A, H17 was the most distantly related haplotype relative to other haplotypes (Fig. 2).

#### Diversity among and within species

Genetic differentiation across all species was high (Table 1, 2) and showed significant phylogeographic structure ( $N_{ST}$  (0.962)  $> G_{ST}$  (0.791),  $P < 0.05$ ). Mantel tests revealed a significant association ( $r^2 = 0.2285$ ,  $P < 0.001$ ) between genetic differentiation and geographic distances between populations over all species. For *C. chengiana* and *C. duclouxiana*, total intraspecific diversity ( $H_T = 0.791$  and  $0.659$ , respectively) across populations was higher, but not substantially, compared with the average within-population diversity ( $H_S = 0.536$  and  $0.443$ , respectively). Within *C. duclouxiana*, interpopulation variation accounted for a high proportion of intraspecific variation (23.3%, Table 2), indicating strong differentiation between populations. The significant phylogeographic structure of *C. chengiana* reflected the presence of two distinct lineages (A and E) within this species that were located in Gansu and Sichuan provinces, respectively. This regional divergence accounted for 44.2% of the total variation recorded within *C. chengiana* (Table 2). Variation within each of these lineages was almost entirely attributed to variation within populations (Table 2, note





that H5 of *C. jiangeensis* was included in the calculations for lineage E). Rare haplotypes were frequent in many populations that were representatives of lineages A or E. In *C. funebris*, two haplotypes were recovered, but no variation was recorded among populations.

### Discussion and Conclusions

Two broad findings emerge from our population-level phylogeographic study of the eight *Cupressus* species that occur in the QTP and adjacent regions. First, there was generally good correspondence between the major plas-

tid DNA lineages/haplotypes identified and the morphological taxonomy of species in the group [12]. Such divergence in plastid DNA sequence between species in this group probably results from their allopatric origins and the maintenance of broadly allopatric distributions following extensive uplifts of the QTP since the Miocene. Second, the amounts and pattern of intraspecific plastid DNA variation recorded differed greatly among the species examined. Species with restricted natural distributions contained low levels of intraspecific diversity relative to those having wide natural distributions. These

**Table 1: Estimates of average genetic diversity within populations ( $H_S$ ), total gene diversity ( $H_T$ ), interpopulation haplotype differentiation ( $G_{ST}$ ), and interpopulation haplotype differentiation taking into account sequence difference ( $N_{ST}$ ) (mean  $\pm$  SE in parentheses)**

Species	$H_S$	$H_T$	$G_{ST}$	$N_{ST}$	$N_{ST} - G_{ST}$
<i>C. chengiana</i>	0.536 (0.079)	0.791 (0.051)	0.322 (0.074)	0.558 (0.056) *	0.236
<i>C. chengiana</i> (lineage E, Sichuan populations)	0.471 (0.136)	0.489 (0.117)	0.037 (0.093)	-0.007 (0.075) <sup>ns</sup>	-0.044
<i>C. chengiana</i> (lineage A, Gansu populations)	0.618(0.052)	0.597(0.060)	-0.035(0.053)	-0.022(0.083) <sup>ns</sup>	0.013
<i>C. funebris</i>	0.254 (0.060)	0.249 (0.050)	-0.021 (0.055)	-0.021 (0.055) <sup>ns</sup>	0
<i>C. duclouxiana</i>	0.443(0.095)	0.659(0.098)	0.328(0.116)	0.488 (0.092) <sup>ns</sup>	0.160
<i>C. gigantea</i>	NC	NC	NC	NC	NC
<i>C. torulosa</i>	NC	NC	NC	NC	NC
<i>C. cashmeriana</i> & <i>C. austrotibetica</i>	NC	NC	NC	NC	NC
<i>C. torulosa</i> s.l. & <i>C. gigantea</i>	0 (0)	0.665(0.039)	1 (NC)	1 (NC)	0
All species	0.190(0.034)	0.908(0.012)	0.791(0.036)	0.962(0.013)*	0.171

\*indicates that  $N_{ST}$  is significantly different from  $G_{ST}$  ( $0.01 < P < 0.05$ ); ns indicates that  $N_{ST}$  is not significantly different from  $G_{ST}$  ( $P > 0.05$ ); NC, not computed due to small sample size; *C. torulosa* s.l. includes *C. torulosa*, *C. cashmeriana* and *C. austrotibetica*

two findings are important to an improved understanding of the importance of geographical isolation on the development of high plant species diversity in the QTP biodiversity hotspot and also the distribution of intraspecific diversity within this region.

#### Allopatric divergence

Phylogenetic relationships among the major plastid DNA lineages detected were largely consistent with those resolved for species in a previous plastid DNA phylogeny constructed by Little [12] based on other plastid DNA sequences and mainly single sample per species. Interestingly, in both phylogenies the Mediterranean *C. sempervirens* and *C. atlantica* were nested within the Asian lineage indicating that they probably originated from this lineage. Phylogenetic relationships resolved in both plastid DNA phylogenies were a little different from those constructed from RAPD or nuclear ITS sequence data [12,28,29]. Because these latter studies sampled only one or very few individuals from each species, it remains unknown whether the nuclear variation reported was species-specific or not. In contrast to the other studies, our investigation is the only one to have examined molecular variation in cypresses at the population level. Our results suggest that two species (*C. austrotibetica* and *C. cashmeriana*) share the same haplotype, which is very similar to the single haplotype detected in *C. torulosa*. However, all three of these species differ in morphology [12]. The fact that two very distinct cpDNA lineages were recovered from northern and southern populations of *C. chengiana*, is of further interest in that plants from these

two regions have sometimes been treated as separate subspecies [13,14]. Thus, our cpDNA data support this division; however, morphological as well as further molecular studies are necessary before a taxonomic revision of this and other Asian cypress species is contemplated.

Our data show that plastid DNA lineage sorting is nearly complete between recognized morphological units of Asian *Cupressus* (Fig. 2). Furthermore, across all populations and species examined, a strong signal for phylogeographic structure is present:  $N_{ST}$  (0.962) was significantly larger than  $G_{ST}$  (0.791,  $P < 0.05$ ) (Table 1), demonstrating that gene-flow is very low between these broadly allopatric species (Fig. 1). According to preliminary calibrations, *C. funebris* diverged from the other lineage/species around 26 million years ago (Mya), while the two major east-west divergences between *C. duclouxiana* and *C. gigantea*, and between *C. chengiana* and the *C. torulosa/C. austrotibetica/C. cashmeriana* lineage, were dated to around 10 Mya and 13 Mya, respectively. Although these estimates of dates of origin should be treated with caution, they correspond well with geological evidence that the QTP was extensively uplifted from the early Miocene to the Pliocene [30-34]. It is likely that these extensive uplifts created fragmentation and isolation of habitats, which promoted repeated allopatric speciation of *Cupressus* in the QTP and adjacent regions. The formation of high mountains separated by deep valleys will have reduced gene flow between *Cupressus* populations that were diverging into different species in different valleys and hastened the fixation of locally spe-

**Table 2: Analysis of molecular variance (AMOVA) for cpDNA haplotypes of cypress species**

Species/lineage	Source of variation	d.f.	SS	VC	Variation (%)	Fixation index
<i>C. chengiana</i> (lineage E, Sichuan populations)	Among populations	5	1.827	0.017	6.8	$F_{ST} = 0.068$
	Within populations	44	10.213	0.232	93.2	
	Total	49	12.040	0.249		
<i>C. chengiana</i> (lineage A, Gansu populations)	Among populations	3	0.925	-0.001	-0.2	$F_{ST} = -0.002$
	Within populations	36	11.325	0.315	100.2	
	Total	39	12.250	0.314		
<i>C. chengiana</i>	Among lineages	1	10.066	0.218	43.9	$F_{CT} = 0.439^*$
	Among populations within groups	8	2.752	0.009	1.8	$F_{SC} = 0.032$
	Within populations	80	21.538	0.269	54.3	$F_{ST} = 0.457^*$
	Total	89	34.356	0.496		
<i>C. funebris</i>	Among populations	10	0.951	-0.004	-3.5	$F_{ST} = -0.035$
	Within populations	78	10.150	0.130	103.5	
	Total	88	11.101	0.126		
<i>C. duclouxiana</i>	Among populations	10	6.707	0.061	23.3	$F_{ST} = 0.233^*$
	Within populations	75	15.142	0.202	76.7	
	Total	85	21.849	0.263		
<i>C. gigantea</i>	Among populations	5	0	0	0	$F_{ST} = 0$
	Within populations	52	0	0	0	
	Total	57	0	0		
<i>C. torulosa</i>	Among populations	11	0.952	0.053	100.0	$F_{ST} = 1.000$
	Within populations	9	0	0	0	
	Total	20	0.952	0.053		
<i>C. cashmeriana</i> & <i>C. austrotibetica</i>	Among populations	7	0	0	0	$F_{ST} = 0$
	Within populations	18	0	0	0	
	Total	25	0	0	0	
<i>C. torulosa</i> s.l. and <i>C. gigantea</i>	Among groups	3	30.659	0.469	97.6	$F_{CT} = 0.976^*$
	Among populations within groups	22	0.955	0.012	2.4	$F_{SC} = 1.000$
	Within populations	80	0	0	0	$F_{ST} = 1.000^*$
	Total	105	31.613	0.481		
All species	Among species	7	100.946	0.319	63.5	$F_{CT} = 0.635^*$
	Among populations within species	58	20.258	0.039	7.7	$F_{SC} = 0.210^*$
	Within populations	305	44.123	0.145	28.8	$F_{ST} = 0.712^*$
	Total	370	165.326	0.502		

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

cific haplotypes. However, allopatric differentiation usually results in the origin of 'paraphyletic' species because of incomplete lineage sorting [6]. Thus, the formation of almost completely mutual 'monophyletic lineages' correlated with morphological units, which we recorded here, is likely to have been brought about by repeated extinctions and/or genetic drift in the isolated habitats [3-5]. A hypothesis of the allopatric origin of each *Cupressus* species in a different valley or valley system is supported by the low level of inter-population differentiation and minimal population substructure observed in most species/lineages except *C. duclouxiana* and *C. chengiana* (Table 2). In these latter two species, a high level of population differentiation was recorded among geographically isolated regions, but very little population divergence within such regions (Fig. 1, Table 2). Two lineages of *C. chengiana* may partly reflect a long history of divergence with the species fragmented into two separate areas in the north and south of western China approximately nine Mya.

The presence of strong geographical isolating barriers in the form of high mountains might also have contributed to rapid species diversification in other plant groups present in the QTP [9,35-38]. For example, in the genus *Ligularia*, most species with distinct morphologies and restricted distributions are thought to have originated through allopatric isolation from a common ancestor within a very short time correlated with the period of extensive uplift of the QTP [36]. Our phylogeographic analysis at the population level of closely related cypress species in the QTP, thus provides further evidence for geographic isolation acting as a driving force of high rates of speciation in this and adjacent regions.

#### Intraspecific diversity

Whereas high levels of plastid DNA diversity were recorded in *C. chengiana* and *C. duclouxiana*, only one or two haplotypes were present in each of *C. gigantea*, *C. austrotibetica*, *C. cashmeriana*, *C. torulosa* and *C. funebris*. This difference in intraspecific diversity is highly correlated with the natural distribution ranges of the sampled species [10-19]. *Cupressus gigantea*, *C. austrotibetica*, *C. cashmeriana* and *C. torulosa* occur sparsely at high altitude in the QTP and the west Himalayas, which have been highly vulnerable to the effects of global climatic change [30]. These species might have been more widespread during periods of more favorable climate in their history. Alternatively or in addition, their restricted distributions might be a consequence of the negative impacts of human activities. Either way, they may have experienced severe population bottlenecks in the relatively recent past which reduced significantly their genetic diversity. In contrast to the other four species, *Cupressus funebris* is distributed in central and eastern

China and adjacent regions where adverse climatic effects should have been less intense. It seems, however, that most natural populations of this species have become extinct in the wild and only a few, comprising trees of large size, survived in south China and north Vietnam [10-19]. However, according to our own field observations, even these populations (for example, the Ruyuan population in Guangdong) (Additional File 1) are likely to have been artificially cultivated because these large trees grow around one ancient temple. The presence of only two haplotypes in this species may be a legacy of its widespread cultivation and eradication from the wild as seems also to be the case for *C. sempervirens* in Europe [39]. In contrast to these Asian cypress species showing low levels of plastid DNA diversity, *C. duclouxiana*, and *C. chengiana* contain high levels of diversity and occur at lower altitude at the eastern margin of the QTP, where they might have experienced fewer bottlenecks and/or founder effects caused by historical climatic change [30]. It is also to be noted that most extant populations of *C. chengiana* are natural, as are populations of *C. duclouxiana* except those cultivated in North Yunnan, and therefore are likely to have been less affected by human activities and subsequent bottleneck effects.

All of the species examined, except for *C. duclouxiana* and *C. funebris*, are listed as endangered [11]. Genetic diversity is critically important for a species' evolutionary potential in the face of environmental change, and loss of genetic diversity is often associated with reduced fitness [40]. The amounts and distribution of genetic diversity within these cypress species are thus important in regard to formulating appropriate plans for conservation purposes. Based on the amounts and distribution of plastid DNA sequence variation within these species, it would seem that for most species, conservation of only a few populations will be sufficient to maintain plastid DNA diversity at current levels. However, the high population differentiation found in *C. duclouxiana* (Table 1) suggests a need to conserve more populations for maintenance of plastid DNA diversity in this species. Similarly, in *C. chengiana* our results make clear that conservation of multiple populations in both Gansu and Sichuan provinces will be necessary to ensure that plastid DNA diversity is maintained in this species.

#### Methods

##### Species and samples

Leaf samples were collected from eight Asian *Cupressus* species that were identified according to the recent taxonomic revision of Little [12]: *C. chengiana* S. Y. Hu, *C. jiangeensis* N. Zhao, *C. cashmeriana* Royle ex Carrière, *C. duclouxiana* Hickel ex Camus, *C. funebris* Endl., *C. torulosa* D. Don, *C. austrotibetica* Silba and *C. gigantea* W. C. Cheng & L. K. Fu. The one Asian species not included

was *C. tonkinensis* Silba, which occurs in Vietnam and across the Vietnam-China border. In total, samples were collected from 371 trees over 66 populations (Additional File 1), which covered most of the natural or cultivated distribution of the species examined. All individuals of *C. austrotibetica* sampled appeared to have been artificially cultivated because they grew around ancient temples, while three populations of *C. duclouxiana* (29, 30 and 31) entirely comprised cultivated trees. Most populations of *C. chengiana* and *C. gigantea* were natural, although a few individuals in the Lixian seemed to have been cultivated. Only one individual of *C. jiangensis* was available for analysis. When more than one sample was collected from a population, samples were taken from trees at least 50 m apart. The latitude, longitude and altitude of localities for most populations sampled were recorded using an Etrex GIS monitor (Garmin, Taiwan). Also surveyed were three cultivated individuals of *C. sempervirens* sampled from England, Italy and China, respectively, and one individual of *C. atlantica*. We sequenced three individuals of the New World cypress species as outgroups because recent phylogenetic analyses suggest that all Old World cypress species comprise a monophyletic lineage that is distinct from the New World cypress lineage and juniper lineage [12][41].

#### DNA extraction, amplification and sequencing

Genomic DNA was isolated from approximately 50 mg of silica-gel dried, leaf-needle material using the CTAB method [42]. Three to five individuals from each of 25 populations sampled across most species were used in an initial screen of plastid DNA polymorphisms. Three different plastid DNA primer pairs were selected for use in this scan, and all revealed polymorphism. Subsequently, we used these three primer pairs to survey all individuals sampled for the *trnD-trnT* [43], *trnS-trnG* [44] and *trnL-trnF* [45] regions, respectively. Polymerase chain reaction (PCR) was performed following the procedure of Zhang et al. [46] in a 25- $\mu$ L volume, containing 10-40 ng plant DNA, 50 mM Tris-HCl (pH 8.0), 1.5 mM MgCl<sub>2</sub>, 0.5 mM dNTPs, 2  $\mu$ M of each primer, and 0.75 unit of Taq polymerase.

PCR products were purified using a TIANquick Midi Purification Kit following the recommended protocol (TIANGEN). Sequencing reactions were performed with the PCR primers described above, to cover the whole PCR segment using ABI Prism BigDye™ Terminator Cycle Sequencing Ready Reaction Kit. Sequencing was performed with BigDye Terminator 3.1 (Applied Biosystems) on an ABI PRISM 3130x1 sequencer using the same primers as used for PCR amplifications. Sequences were aligned with CLUSTAL X [47] and double-checked manually. A matrix of combined sequences was constructed for the 371 individuals examined, and showed that 30 dif-

ferent cpDNA sequences (haplotypes) were present. All sequences have been deposited in the GenBank databases under accession numbers [HM008329-HM008379](#).

#### Data analysis

We used PAUP 4.0 [48] to perform maximum parsimony and neighbour joining analyses to examine phylogenetic relationships among cpDNA haplotypes (Additional File 3). All indels were coded as binary characters (0 or 1) using the GAPCODER program [49] and those with further indels were sequentially assumed as multiple evolutionary events. Maximum parsimony analyses (equally weighted characters and nucleotide transformations) involved a heuristic search strategy with 100 replicates of random addition of sequences, in combination with ACCTRAN character optimization and MULTIPARS+TBR branch-swapping and STEEPEST DESCENT options on. Bootstrap values for MP analyses (1000 replicates) were used to assess the relative support for monophyletic groups [50].

We used BEAST (under the common model GTR) [51] to estimate the mutation rate of *Cupressus* and the possible divergence timescales of the main cpDNA lineages across the eight Asian cypress species. We edited two cpDNA (*trnD-trnT*, *trnS-trnG* and *trnL-trnF*) datasets respectively. The first comprised three *Cupressus* species, three *Juniperus* species, three *Calitropsis* species, one species for *Platycladus*, *Calocedrus*, *Chamaecyparis* and *Thuja*. The other included the most common haplotypes of the *Cupressus* species and three *Calitropsis* species. For analyses of both datasets, we excluded all indels and only nucleotide mutations were considered. A strict molecular clock assumption was rejected for both datasets because of significant rate heterogeneity among lineages. We used BEAST version 1.5.3 employing a Bayesian Markov chain Monte Carlo (MCMC) chain and co-estimated topology to estimate the possible mutation rate of *Cupressus* and divergence timescales across the Asian cypress species. Under the GTR model of nucleotide substitution with a gamma distribution and four rate categories, three tree prior models (Coalescent: constant size, exponential growth; and Speciation: Yule process) were implemented with rate variation across branches assumed to be uncorrelated and log-normally distributed. The model that yielded the highest posterior probability estimates was employed to perform a final analysis. For all analyses, posterior distributions of parameters were approximated using two independent MCMC analyses of 100,000,000 generations with 20% burn-in. The program Tracer v1.5 [51] was used to check Effective Sample Size (ESS) and the program TreeAnnotator 1.5.3 (part of the BEAST 1.5.3 package) was used to combine all samples and to converge and/or summarize output results. A tree with ages for each node and their



95% HPD were displayed and modified in FigTree v1.3.1 [51]. According to the fossil calibrations of the first dataset, a mutation rate of cpDNA ( $3.2 \times 10^{-10}$  substitutions per site per year) was obtained (Additional File 4). This mutation rate was used to calibrate the interspecific or interlineage divergences across the cypress species (Additional File 4).

An examination of interrelationships between haplotypes within lineages recovered in phylogenetic trees was conducted using TCS 1.21 [52]. TCS was run with the default parsimony connection limit of 95%. We assumed that all coded indels and mutations evolved at an equal rate. Estimates of unbiased genetic diversity ( $H_E$ ) [equivalent to expected heterozygosity for diploid data] were calculated as  $\hat{s} = n/(n-1) (1 - \sum p_i^2)$ , in which  $n$  is the number of samples in the population,  $k$  is the number of haplotypes, and  $p$  is the population frequency of the  $i$ th haplotype [53]. Where there were insufficient numbers of individuals in a population, multiple nearby populations were combined into a larger one for analysis (for example, populations 43-46, each comprising a single tree, were treated together as one population). Average gene diversity within populations ( $H_S$ ), total gene diversity ( $H_T$ ) and two measures of population differentiation,  $G_{ST}$  (coefficient of genetic variation over all populations [54]) and  $N_{ST}$  (equivalent coefficient taking into account sequence similarities between haplotypes) were calculated using the program PERMUT (available at <http://www.pierroton.inra.fr/genetics/labo/Software/>) [55]. Significant phylogeographic structure was inferred by testing whether  $N_{ST}$  was significantly greater than  $G_{ST}$  using PERMUT with 1000 permutations. Genetic structure was further examined by analysis of molecular variance (AMOVA, [56]) as implemented in Arlequin version 3.01 [57]. The significance of isolation by distance between populations was tested using a Mantel test with 1000 random permutations on matrices of pairwise population  $F_{ST}$  values and the natural logarithm of geographical distances [58]. Pairwise  $F_{ST}$  values between populations were estimated using Arlequin while geographical distances between populations were calculated with the help of the program at <http://www.indo.com/distance/> (e.g. [59]).

## Additional material

**Additional File 1** Locations of sites and voucher specimens of eight Asian *Cupressus* species surveyed for cpDNA variation.

**Additional File 2** Haplotype frequencies and genetic diversity ( $H_E$ ) of populations of eight Asian *Cupressus* species surveyed for plastid DNA variation.

**Additional File 3** Variable sites for aligned sequences of three plastid DNA fragments among 30 haplotypes resolved in eight Asian *Cupressus* species (-, one indel). Sequences are numbered from the 5' to the 3' end in each region.

### Additional File 4 Estimates of mutation rate for *Cupressus* and dates of divergence between the Asian species/lineage estimated by BEAST.

#### Authors' contributions

JQL designed the study. JQL, GM and ZXCR collected materials. TTX and GLW finished molecular experiments. TTX, KSM and FKD analyzed data. JQL, RJA, RIM and TTX wrote the manuscript. All authors read and approved the final manuscript.

#### Acknowledgements

This research was supported by grants from the National Natural Science Foundation of China (30725004, 30430560) and Key Innovation Project of Ministry of Education of China to JQL, and a Royal Society-NSA China International Joint Project award 2010/R4 to RJA and JQL.

#### Author Details

<sup>1</sup>Division of Molecular Ecology, Key Laboratory of Arid and Grassland Ecology, School of Life Science, Lanzhou University, Lanzhou, 730000, China, <sup>2</sup>School of Biology, Mitchell Building, University of St Andrews, St Andrews, Fife KY16 9TH, UK, <sup>3</sup>Institute of Molecular Plant Sciences, The University of Edinburgh, Daniel Rutherford Building, King's Buildings, Mayfield Road, Edinburgh EH9 3JH, UK, <sup>4</sup>Department of Life Science, Tibet University, Lasha, Tibet, China and <sup>5</sup>Faculty of Geography, University of Marburg, Deutschhastr.10 35032 Marburg, Germany

Received: 5 August 2009 Accepted: 22 June 2010

Published: 22 June 2010

#### References

1. Mayr E: **Systematics and the Origin of Species**. New York: Columbia University Press; 1942.
2. Sobel JM, Chen GF, Watt LR, Schemske DW: **The biology of speciation**. *Evolution* 2010, **64**:295-315.
3. Coyne JA, Orr HA: **Speciation**. Massachusetts: Sinauer Associates Inc; 2004.
4. Orr HA: **Dobzhansky, Bateson, and the genetics of speciation**. *Genetics* 1996, **144**:1331-1335.
5. Jaramillo-Correa JP, Aguirre-Planter E, Khasa DP, Eguiarte LE, Piñero D, Furnier GR, Bousquet J: **Ancestry and divergence of subtropical montane forest isolates: molecular biogeography of the genus *Abies* (Pinaceae) in southern México and Guatemala**. *Mol Ecol* 2008, **17**:2476-2490.
6. Avise JC: **Phylogeography: The History and Formation of Species**. Cambridge, Massachusetts: Harvard University Press; 2000.
7. Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J: **Biodiversity hotspots for conservation priorities**. *Nature* 2000, **403**:853-858.
8. Yang FS, Li YF, Ding X, Wang XQ: **Extensive population expansion of *Pedicularis longiflora* (Orobanchaceae) on the Qinghai-Tibetan Plateau and its correlation with the Quaternary climate change**. *Mol Ecol* 2008, **17**:5135-5145.
9. Wang LY, Abbott RJ, Zheng W, Chen P, Wang YJ, Liu JQ: **History and evolution of alpine plants endemic to the Qinghai-Tibetan Plateau: *Aconitum gymnandrum* (Ranunculaceae)**. *Mol Ecol* 2009, **18**:709-721.
10. Farjon A: **A Monograph of Cupressaceae and Sciadopitya**. *Kew: Royal Botanic Gardens* 2005.
11. Fu LG, Jin JM: **China Plant Red Data Book--Rare and Endangered Plants**. Beijing, China: Science Press; 1992.
12. Little DP: **Evolution and circumscription of the True Cupresses (Cupressaceae: *Cupressus*)**. *Syst Bot* 2006, **31**:461-480.
13. Silba J: **The Trans-Pacific relationship of *Cupressus* in India and North America**. *J Int Conifer Preserv Soc* 1994, **1**:1-28.
14. Silba J: **A monograph of the genus *Cupressus* L.** *J Int Conifer Preserv Soc* 1998, **5**:1-100.
15. Silba J: **A monograph of the genus *Cupressus* L. in the twenty-first century**. *J Int Conifer Preserv Soc* 2005, **12**:30-103.
16. Fu LG, Yu YF, Farjon A: **Cupressaceae**. In *Flora of China* St. Louis, USA and Beijing, China: Science Press and Missouri Botanical Garden Press; 1999:65-67.
17. Zhao N: **Species nova generis Cupressi**. *Acta Phytotax Sin* 1980, **18**:210.

18. Silba J: Revised generic concepts of *Cupressus* L. (Cupressaceae). *Phytologia* 1981, **49**:394.
19. Silba J: A new species of *Cupressus* L. from Tibet (Cupressaceae). *Phytologia* 1988, **65**:333-336.
20. Neale DB, Marshall KA, Harry DE: Inheritance of chloroplast and mitochondrial DNA in incense-cedar (*Calocedrus decurrens*). *Can J For Res* 1991, **21**:717-720.
21. Mogensen HL: The hows and whys of cytoplasmic inheritance in seed plants. *Am J Bot* 1996, **83**:383-404.
22. Kondo T, Sumura Y, Kawahara T, Okamura M: Inheritance of chloroplast and mitochondrial DNA in interspecific hybrids of *Chamaecyparis* spp. *Breed Sci* 1998, **78**:177-179.
23. Ennos RA, Sinclair WT, Hu X-S, Langdon A: Using organelle markers to elucidate the history, ecology and evolution of plant populations. In *Molecular Systematics and Plant Evolution* Edited by: Hollingsworth PM, Bateman RM, Gornall RJ. London: Taylor and Francis; 1999:1-19.
24. Petit RJ, Duménil J, Fineschi S, Hampe A, Salvini D, Vendramin GG: Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Mol Ecol* 2005, **14**:689-701.
25. Currat M, Ruedi M, Petit RJ, Excoffier L: The hidden side of invasions: massive introgression by local genes. *Evolution* 2008, **62**:1908-1920.
26. Du KF, Petit RJ, Liu JQ: More introgression with less gene flow: Chloroplast vs. Mitochondrial DNA in the *Picea asperata* complex in China, and comparison with other conifers. *Mol Ecol* 2009, **18**:1396-1407.
27. Little DP, Stevenson D: A comparison of algorithms for the identification of specimens using DNA barcodes: examples from gymnosperms. *Cladistics* 2007, **23**:1-21.
28. Rushforth K, Adams RP, Zhong M, XQiang M, Pandley RN: Variation among *Cupressus* species from the eastern hemisphere based on Random Amplified Polymorphic DNAs (RAPDs). *Biochem Syst Ecol* 2003, **31**:17-24.
29. Xiang Q, Li J: Derivation of *Xanthocyparis* and *Juniperus* from within *Cupressus*: evidence from sequences of nrDNA internal transcribed spacer region. *Harvard Pap Bot* 2005, **9**:375-382.
30. Shi YF, Li JJ, Li BY: Uplift and Environmental Changes of Qinghai-Tibetan Plateau in the Late Cenozoic. Guangzhou, China: Guangdong Science and Technology Press; 1998.
31. Harrison TM, Copeland P, Kidd WSF, Yin A: Raising Tibet. *Science* 1992, **255**:1663-1670.
32. Spicer R, Harris N, Widdowson M, Herman A, Guo S, Valdes P: Constant elevation of southern Tibet over the past 15 million years. *Nature* 2003, **421**:622-624.
33. Guo ZT, Ruddiman WF, Hao QZ, Wu HB, Qiao YS, Zhu RX, Peng SZ, Wei JJ, Yuan BY, Liu TS: Onset of Asian desertification by 22 Myr ago inferred from loess deposits in China. *Nature* 2002, **416**:159-163.
34. An ZS, Kutzbach JE, Prell WL, Port SC: Evolution of Asian monsoons and phased uplift of the Himalayan-Tibetan Plateau since late Miocene times. *Nature* 2001, **411**:62-66.
35. Wang LY, Ikeda H, Liu TL, Wang YJ, Liu JQ: Repeated range Expansion and Glacial Endurance of *Potentilla glabra* (Rosaceae) in the Qinghai-Tibetan Plateau. *J Integr Plant Biol* 2009, **51**:698-706.
36. Liu JQ, Wang YJ, Wang AL, Ohba H, Abbott R: Radiation and diversification of *Ligularia-Cremnathodium-Parasenecio* triggered by uplifts of the Qinghai-Tibetan Plateau. *Mol Phylogenet Evol* 2006, **38**:31-49.
37. Wang YJ, Susanna A, von Raab-Straube E, Milne R, Liu JQ: Island-like radiation of *Saussurea* (Asteraceae: Cardueae) triggered by uplifts of the Qinghai-Tibetan Plateau. *Biol J Linn Soc* 2009, **97**:893-903.
38. Liu JQ, Gao TG, Chen ZD, Lu AM: Molecular phylogeny and biogeography of the Qinghai-Tibet Plateau endemic *Nannoglottis* (Asteraceae). *Mol Phylogenet Evol* 2002, **23**:307-325.
39. Bagnoli F, Vendramin GG, Buonamici A, Doullis AG, Gonzálezmartínez SC, Laporta N, Magri D, Raddi P, Sebastiani F, Fineschi S: Is *Cupressus sempervirens* native in Italy? An answer from genetic and palaeobotanical data. *Mol Ecol* 2009, **18**:2276-2286.
40. Guerrant EO: Genetic and demographic considerations in the sampling and reintroduction of rare plants. In *Conservation Biology: The Theory and Practice of Nature Conservation, Preservation and Management Volume 1992*. Edited by: Fiedler PL, Jain SK. New York: Chapman & Hall :321-344.
41. Mao KS, Hao G, Liu JQ, Adams RP, Milne RI: Diversification and biogeography of *Juniperus* (Cupressaceae): variable diversification rates and multiple intercontinental dispersals. *New Phytol* 2010 in press.
42. Doyle JJ, Doyle JL: A rapid DNA isolation procedure for small quantities of fresh leaf material. *Phytochem Bull* 1987, **19**:11-15.
43. Demesure B, Sodzi N, Petit RJ: A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Mol Ecol* 1995, **4**:129-131.
44. Hamilton MB: Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Mol Ecol* 1999, **8**:521-523.
45. Taberlet P, Gielly L, Pautou G, Bouvet J: Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol Biol* 1991, **17**:1105-1109.
46. Zhang Q, Chiang TY, George M, Liu JQ, Abbott RJ: Phylogeography of the Qinghai-Tibetan Plateau endemic *Juniperus przewalskii* (Cupressaceae) inferred from chloroplast DNA sequence variation. *Mol Ecol* 2005, **14**:3513-3524.
47. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997, **25**:4876-4882.
48. Swofford DL: PAUP\* *Phylogenetic Analysis Using Parsimony (\* and Other Methods)*, Version 4. Sunderland, Massachusetts: Massachusetts Sinauer Associates; 2000.
49. Young ND, Healy J: GapCoder automates the use of indel characters in phylogenetic analysis. *BMC Bioinformatics* 2003, **4**:1-6.
50. Felsenstein J: Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985, **39**:783-791.
51. Drummond AJ, Rambaut A: BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007, **7**:214.
52. Clement M, Posada D, Crandall KA: TCS: a computer program to estimate gene genealogies. *Mol Ecol* 2000, **9**:1657-1660.
53. Nei M: *Molecular Evolutionary Genetics*. New York: Columbia University Press; 1987.
54. Nei M: Analysis of gene diversity in subdivided populations. *PNAS* 1973, **70**:3321-3323.
55. Pons O, Petit RJ: Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* 1996, **144**:1237-1245.
56. Excoffier L, Smouse PE, Quattro JM: Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 1992, **131**:479-491.
57. Excoffier L, Laval G, Schneider S: ARLEQUIN (version 3.01): an integrated software package for population genetic data analysis. *Evol Bioinform Online* 2005, **1**:47-50.
58. Rousset F: Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics* 1997, **145**:1219-1228.
59. Chen KM, Abbott RJ, Milne RI, Tian XM, Liu JQ: Phylogeography of *Pinus tabulaeformis* Carr. (Pinaceae), a dominant species of coniferous forest in northern China. *Mol Ecol* 2008, **17**:4276-4288.

doi: 10.1186/1471-2148-10-194

Cite this article as: Xu et al., Phylogeography and allopatric divergence of cypress species (*Cupressus* L.) in the Qinghai-Tibetan Plateau and adjacent regions *BMC Evolutionary Biology* 2010, **10**:194