

RESEARCH ARTICLE

Open Access



# Phylogeographic patterns of the desert poplar in Northwest China shaped by both geology and climatic oscillations

Yan-Fei Zeng<sup>1</sup> , Jian-Guo Zhang<sup>1,2,3\*</sup>, Bawerjan Abuduhamiti<sup>4</sup>, Wen-Ting Wang<sup>5</sup> and Zhi-Qing Jia<sup>6</sup>

## Abstract

**Background:** The effects of historical geology and climatic events on the evolution of plants around the Qinghai-Tibetan Plateau region have been at the center of debate for years. To identify the influence of the uplift of the Tianshan Mountains and/or climatic oscillations on the evolution of plants in arid northwest China, we investigated the phylogeography of the Euphrates poplar (*Populus euphratica*) using chloroplast DNA (cpDNA) sequences and nuclear microsatellites, and estimated its historical distribution using Ecological Niche Modeling (ENM).

**Results:** We found that the Euphrates poplar differed from another desert poplar, *P. pruinosa*, in both nuclear and chloroplast DNA. The low clonal diversity in both populations reflected the low regeneration rate by seed/seedlings in many locations. Both cpDNA and nuclear markers demonstrated a clear divergence between the Euphrates poplar populations from northern and southern Xinjiang regions. The divergence time was estimated to be early Pleistocene based on cpDNA, and late Pleistocene using an Approximate Bayesian Computation analysis based on microsatellites. Estimated gene flow was low between these two regions, and the limited gene flow occurred mainly via dispersal from eastern regions. ENM analysis supported a wider distribution of the Euphrates poplar at 3 Ma, but a more constricted distribution during both the glacial period and the interglacial period.

**Conclusions:** These results indicate that the deformation of the Tianshan Mountains has impeded gene flow of the Euphrates poplar populations from northern and southern Xinjiang, and the distribution constriction due to climatic oscillations further accelerated the divergence of populations from these regions. To protect the desert poplars, more effort is needed to encourage seed germination and seedling establishment, and to conserve endemic gene resources in the northern Xinjiang region.

**Keywords:** Arid northwest China, Climatic oscillations, Desert poplar, Ecological Niche Modeling, Phylogeography, Tianshan Mountains

## Background

In eastern Asia, the effect of geological dynamics, e.g., the uplift of the Qinghai-Tibetan Plateau (QTP) and Quaternary climatic oscillations, on species diversification has been at the center of debate for several years [1–4]. A lot of phylogeography studies have focused on the speciation and population divergence of organisms

in the Hengduan Mountains and adjacent regions in southwest China, which harbors one of the world's major plant diversity hotspots [3]. Studies have found that orogeny creates conditions favoring rapid in situ speciation of resident lineages [4, 5], and the lowland refugia for plants throughout climatic oscillations in the Quaternary further maintain high biodiversity in these regions [2]. Recently, a growing number of studies have begun to pay attention to plants near the northern edge of the QTP in the arid northwest area of China [6]. However, the possible roles of geology and climatic oscillations in facilitating the population divergence and species diversification of plants in arid northwest China are still poorly understood.

\* Correspondence: [zhangjg@caf.ac.cn](mailto:zhangjg@caf.ac.cn)

<sup>1</sup>State Key Laboratory of Tree Genetics and Breeding, Research Institute of Forestry, Chinese Academy of Forestry, NO. 1 Dongxiaofu, Xiangshan road, Haidian district, Beijing 100091, China

<sup>2</sup>Collaborative Innovation Center of Sustainable Forestry in Southern China, Nanjing Forestry University, No. 159 Longpan road, Nanjing 210037, China  
Full list of author information is available at the end of the article



Northwest China is located in central Asia and is arid or semi-arid, with a cold and dry continental climate. The progressive extension of the uplift of the QTP was associated with the orogeny of the Tianshan Mountains, which achieved a significant elevation during the Miocene [3, 7]. The rise of the Tianshan Mountains and the Higher Himalayas massively altered air circulation. Meanwhile, a worldwide cooling has occurred since the Middle Miocene climate optimum. These together have caused the progressive aridification of central Asia [8]; see [3] for more references. Other factors, such as changes in the global ice volume and the final disappearance of the Tethys Sea in Asia (late Miocene and early Pliocene), and the intensification of the East Asian summer and winter monsoons (EASM and EAWM) also contributed to the aridification of central Asia [9, 10]. Currently, the deserts of northwest China, including the four largest deserts (Taklamakan, Guerbantonggute, Badain Jaran, and Tengger), together comprise the world's largest mid-latitude, temperate, continental interior desert region [11, 12].

The flora of arid northwest China constitutes mainly Tethys coastal xerophytes [13]. The Tianshan Mountains, running roughly E-W for about 2500 km, are located between the Tarim and Dzungarian basins and separate the Taklamakan and Guerbantonggute deserts, and therefore may have profoundly affected the genetic structure and distribution patterns of desert plants in northwest China. For example, these mountains are suggested to have triggered the separation of different *Aconitum nemoru* lineages [14]. Although the major Quaternary glaciations were absent in arid northwest China, the evolution of the biota there have been impacted by significant climatic oscillations [6]. The further development of the East Asian monsoon system, especially the increasing variability and strengthening of the EAWM, may have influenced the onset of major Northern Hemisphere glaciations after 2.6 Mya, and accelerated the further aridification of central Asia during the Pliocene and Pleistocene [10]. Previous phylogeographical studies have found that desert expansion caused habitat fragmentation and aridification during the last glacial maximum (LGM), which promoted the diversification and speciation of desert plants [15], see [6] for more references. Although a number of recent studies have examined the effects of a particular process in a limited geographic region [6], few have considered the relative importance of geological and climatic dynamics over a wide area. Related studies may facilitate our understanding the roles of geology and climatic oscillations in driving population divergence of plants in this arid zone, and comparing them with those in the diversity hotspots regions in southwest China, which would give an insight into the emergence and maintenance of biodiversity.

Two closely related desert poplars, the Euphrates poplar (*Populus euphratica* Olivier) and *P. pruinosa* Schrenk, are the only tree species that have established in the world's largest shifting-sand desert, the Taklamakan Desert [16]. *P. euphratica* is naturally distributed in western China and adjacent Middle-Eastern countries. More than 61% of *P. euphratica* forests occur in China, of which about 91.1% are located in Xinjiang province [16]. *P. pruinosa* has a more restricted distribution than *P. euphratica*, occurring in alluvial oasis communities in northwest China, Kazakstan, Tajikistan, Turkmenistan, and Uzbekistan [17]. Recent studies suggest that gene flow and Pleistocene climate oscillations might have triggered this speciation [17, 18]. Populations of *P. euphratica* and *P. pruinosa* are endangered as a consequence of intensive water use and damming in northwest China [19, 20]. As a typical species of Tethys, the Euphrates poplar might have existed before the uplift of the Tianshan Mountains, and therefore it is a suitable species in which to study the influence of geological and climatic dynamics on species divergence in arid northwest China.

In this study, we used both chloroplast and nuclear markers to explore the phylogeographical pattern of these two desert poplars in northwest China, with a particular focus on the widespread Euphrates poplar, and used Ecological Niche Modeling (ENM) to estimate their potential historical distributions, and finally identify how the uplift of the Tianshan Mountains and Quaternary climatic oscillations influenced the current population structure of desert plants. More specifically, the aim was to characterize: (1) the pattern of genetic diversity of the two desert poplars in northwest China; (2) the divergence and gene flow between Euphrates poplar populations to the north and the south of the Tianshan Mountains; and (3) the historical population demography of Euphrates poplar in response to climatic oscillations.

## Methods

### Sampling and DNA extraction

Leaf samples from 552 *P. euphratica* individuals were collected from 33 natural populations (12 from northern Xinjiang, 14 from southern Xinjiang, and seven from Qinghai province, Gansu province, and Inner Mongolia province (the 'QGM' region)), which covers the whole range of this species in northwest China (See Additional file 1: Figure S1 and Table S1). Leaf samples of 102 *P. pruinosa* individuals were collected from five natural populations in Xinjiang province, of which four co-occurred with *P. euphratica*. All sampled individuals were at least 20 m apart from each other in any given population. Leaf tissues were dried with silica gel and taken to the laboratory. Total genomic DNA was extracted from 25 mg of leaf tissue from each tree and

purified using a Plant Genomic DNA Extraction Kit (Tiangen, Beijing, China). In addition, DNA samples from three *P. ilicifolia* supplied by the Royal Botanic Gardens (<http://www.kew.org/>) were used for chloroplast DNA (cpDNA) sequence analysis. All collection of specimens used in our study complied with related institutional, national, and international guidelines.

### Microsatellite marker procedure

The samples were screened for variation at 17 nuclear microsatellite loci that were either supplied by the International Populus Genome Consortium from genome sequence of *P. trichocarpa* [21] or developed for *P. euphratica* [22]; see Additional file 1: Table S2 for primer details. We used an economic method suggested by Schuelke [23] for fluorescent dye labeling of the polymerase chain reaction (PCR) fragments. This was performed using a three-primer protocol including unlabeled M13-tagged forward and unlabeled/untagged reverse primers for each marker and a third ‘universal’ M13-primer labelled with one of the fluorescent dyes, 6-FAM, HEX, or TAMRA (Sangon, Shanghai, China). PCR amplifications were performed according to Schuelke [23]. Microsatellite genotypes were resolved on an ABI 3130XL automated sequencer (Life Technologies, Foster City, CA, USA), for which allele sizing was performed using the GENEMAPPER software version 4.0 (Life Technologies), with the Liz 500 (Life Technologies) as an internal standard.

### Chloroplast sequence procedure

The chloroplast *trnK* region was amplified using the primers *trnK*-1682F (5′- GGGTTGCCCGGGACTCGAA C-3′) and *trnK*-4292R (5′- TGGGTTGCTAACTCAAT GG-3′), which were improved from the Demesure et al. [24] original, according to the *P. trichocarpa* chloroplast complete genome (GenBank accession numbers NC\_009143) [21]. A PCR was performed in a 25-μL volume, following the methods described by Zeng et al. [25]. Sequencing reactions were performed with the PCR primers and two more in-between primers, *matK*-2332F (5′- ACTAATGGGATGTCCTACTG-3′) and *matK*-374 1R (5′- GATTTCTAGTCACCTATTAC-3′), to cover the whole PCR segment, using an ABI Prism BigDye™ terminator cycle sequencing ready reaction kit (Life Technologies). The reaction mixtures were analyzed on an ABI 3130xl automated sequencer (Life Technologies).

### Microsatellite data analysis

#### Genetic diversity and differentiation

Clonal reproduction is common in *P. euphratica* and *P. pruinosa* [26]. To identify the clonal lineage of each population, multilocus genotype assignments were conducted using the GenoType software [27], by calculating

a pairwise distance matrix under an infinite allele mutation model. A threshold of two was used to define the same clonal lineage after drawing a frequency distribution of the values of all pairwise comparisons (See Additional file 1: Figure S2).

Clonal diversity was estimated as:

$$R = \frac{G-1}{N-1}$$

with  $G$  representing the number of multilocus genotypes (MLGs) or multilocus lineages (when taking into account possible somatic mutations or scoring errors) discriminated in the sample and  $N$  representing the number of sampled ramets.

After the identification of ramets belonging to the same genets, replicates were removed from the data set to perform a subsequent analysis. The number of alleles, allele frequencies, and observed and expected heterozygosity ( $H_O$  and  $H_E$ ), were calculated for each microsatellite locus using the FSTAT 2.9.3 program [28]. With this program, genetic diversity in terms of total allele numbers ( $A$ ),  $H_O$ ,  $H_E$ , and the inbreeding coefficient ( $F_{IS}$ ) were calculated for each population across all microsatellite loci. We used a permutation procedure (1000 permutations) to test whether a particular estimate of the overall  $F_{IS}$  was significantly different from 0. For *P. euphratica*, the regional genetic diversity (northern Xinjiang, southern Xinjiang, and QGM regions) was also calculated using values of  $H_O$ ,  $H_S$  (average of  $H_E$  for subpopulations), and  $F_{IS}$ . Allelic richness ( $A_R$ ) and private allelic richness ( $P_A$ ) for each of the three regions were calculated by standardization for 25 individuals, using the hierarchical sampling method as executed in *hp\_rare* 1.0 [29]. For both species, population differentiation among populations and among regions were estimated by  $F_{ST}$  with FSTAT [28].

#### Bayesian cluster analysis

We used a Bayesian model-based clustering method implemented in the program STRUCTURE version 2.3 [30–32] to first identify *P. euphratica* or *P. pruinosa* ancestry for all individuals, and then to detect the population structure of *P. euphratica* for individuals only with *P. euphratica* morphology. Based on the LOCPRIOR model described by Hubisz et al. [33], the program was run based on genet MLG data, with a correlated allele frequency model ( $F$ -model) and an admixed origin of populations. After an initial test, where we varied the burn-in and run length, the burn-in was set to 500,000 with 1,000,000 additional cycles. For both data sets, 20 replicate runs were conducted for each value of  $K$  from one to ten. The final posterior probability of  $K$ ,  $\Pr(X|K)$ , and  $\Delta K$ , where the modal value of the distribution is

located at the real  $K$  [34], were both used as an indication of the most likely number of clusters. For graphic visualization of the STRUCTURE results, we used DISSTRUCT [35].

### Population history inference

To identify the possible population history of *P. euphratica*, five alternative scenarios for three population groups, northern and southern Xinjiang, and the QGM region, were tested using the Approximate Bayesian Computer procedure [36] as performed in DIYABC v.2.1.0 [37]. Scenarios 1, 2, and 3 tested a ‘colonization’ event from the QGM region, and northern and southern Xinjiang, respectively. Scenario 4 tested a split of the three groups at the same time. In Scenario 5, populations from the QGM region were created by an admixture of two separated gene pools from northern and southern Xinjiang. We set a varied ancestry population size before the colonization or split of the populations (See Additional file 1: Figure S3).

One million simulations were run for each scenario. Prior parameter distributions for population sizes and time frames (measured in generations) were set as following: uniform (100; 100,000) for both current and ancestral effective population sizes, uniform (10; 10,000) for divergence times  $t_1$ , uniform (10; 50,000) for  $t_2$  (with  $t_2 > t_1$ ), and uniform (0.001; 0.999) for the admixture rate. *P. euphratica* begins to reproduce at the age of 8–10 years and reaches flourishing period at the age of 15–30 years [38]; we thus considered 20 years to represent a reasonable generation time.

### Historical and contemporary gene flow

To estimate the gene flow pattern among regional population groups of *P. euphratica* (northern and southern Xinjiang, and the QGM region), we used the coalescent approach implemented in Migrate-n version 3.6 [39]. This program calculates maximum likelihood (ML) estimates for both effective population size ( $\theta = 4N_e\mu$ , where  $\mu$  = mutation rate) and historical migration rates ( $M = m/\mu$ , where  $m$  = migration rate) between pairs of population groups [40]. Analyses were run for five replicates using a Brownian motion mutation model, with constant mutation rates and starting parameters based on  $F_{ST}$  calculations. For each replicate, we used 10 short chains (10,000 trees) and three long chains (200,000) with 100,000 trees discarded as an initial ‘burn-in’ and a static heating scheme at four temperatures (1, 1.5, 3, and 1000,000). To avoid the confounding effects of the differences in sample size on the estimate of gene flow, the software randomly picked 40 individuals from each group.

To estimate contemporary gene flow ( $m$ ) between the above three regional population groups of *P. euphratica*,

we used the software BayesAss 3.0 [41], which carries assignment tests in a Bayesian framework and a Markov coupled Markov chain (MCMC). After a burn-in of  $2 \times 10^6$ , we ran the analyses for  $10^7$  iterations, with a sampling frequency of  $10^3$ . Delta values for the migration ( $m$ ), allele frequencies ( $a$ ) and inbreeding ( $f$ ) switching proposals were adjusted so that the accepted numbers of changes were 30–50% of the total number of iterations. The delta values for  $a$ ,  $m$ , and  $f$  were 0.2, 0.1, and 0.2, respectively. We performed 20 runs (each with a different seed) and selected the one with the lowest deviance for further analyses [42].

### CpDNA sequence analyses

Sequences of *trnK* were aligned using ClustalX ver. 1.81 (Thompson et al., 1997) with subsequent manual adjustments. A matrix of *trnK* sequences was constructed for the 286 trees that we examined, and different cpDNA sequences were identified as haplotypes. The sequences for those haplotypes have been deposited in GenBank under accession numbers KY002202–KY002230. Average chloroplast gene diversity within populations ( $H_S$ ) and total gene diversity ( $H_T$ ) were calculated using the program ARLEQUIN version 3.1 [43] for each species and each region of *P. euphratica*, respectively. The measure of population differentiation,  $G_{ST}$ , was estimated as  $(H_T - H_S)/H_T$ . Relationships between haplotypes were examined via a haplotype network, which was constructed using the computer program Network version 4.5.1.6 [44]. Phylogenetic relationships and divergence times among haplotypes were estimated using Bayesian inference methods implemented in BEAST v.1.8.2 [45], using a sample with each of *P. alba* and *P. laurifolia* as outgroups and the substitution rate of the chloroplast sequence estimated in *P. balsamifera* and *P. trichocarpa* ( $\mu = 3.46 \times 10^{-10} \text{ s}^{-1} \text{ y}^{-1}$ ; [46]). The General Time Reversible (GTR) nucleotide substitution model was used in the program. A normal prior probability distribution was used to accommodate the uncertainty of the prior knowledge. We sampled all parameters once every 1000 steps from  $10^7$  MCMC steps, with the first 25% of samples discarded as the burn-in. The consensus trees were generated by TreeAnnotator v.1.8.2 [45].

### Ecological niche modeling

Species distribution models for *P. euphratica* and *P. pruinosa* were generated using MAXENT 3.3.3e [47], to predict species occurrence under present-day, LGM, last interglacial (LIG), and three million years ago (Pliocene) conditions. In addition to our sample locations, the distribution records for the Euphrates poplar were sourced from the Chinese Virtual Herbarium (<http://www.cvh.ac.cn/>), Global Biodiversity Information Facility (<http://www.gbif.org/>), and previously published

papers [17, 38, 48]. The locations where species occurred that were within 24 km of one another (12 arc-min) were removed to reduce the effects of spatial autocorrelation in climate variables. The ecological layers for the current climate were obtained from the WorldClim database (<http://www.worldclim.org/version1>) [49]. For the LGM prediction, data were taken from general circulation model simulations using the Community Climate System Model (CCSM) [50] and the Model for Interdisciplinary Research on Climate (MIROC) [51]. The LIG distributions for these two species were predicted using the climatic model developed by Otto-Bliesner et al. [52]. For the Pliocene prediction, data provided by Lunt et al. [53] were used.

For *P. euphratica*, the following five uncorrelated and biologically significant bioclimatic variables were selected as predictors: (1) mean diurnal range (mean of monthly maximum temperature - minimum temperature), (2) minimum temperature of the coldest month, (3) mean temperature of the wettest quarter, (4) precipitation of the driest quarter, and (5) precipitation of the warmest quarter. For *P. pruinosa*, (1) annual mean temperature, (2) temperature annual range (maximum temperature of warmest month - minimum temperature of coldest month), (3) precipitation of wettest month, (4) precipitation of driest month, and (5) precipitation of coldest quarter were used. After a model testing with 25% of the data, model validation was then performed 20 independent replicates using default settings. For each run, the area under the receiver operating characteristic curve was calculated as an indicator of the accuracy of model prediction. The ENM was first analyzed based on worldwide records of the Euphrates poplar, and then analyzed based on records in China only to specify the suitability of its estimation in China.

## Results

### Microsatellite variation

The 17 microsatellite loci yielded a total of 228 alleles (2–28 per locus) from our sample of 673 individuals (See Additional file 1: Table S3), of which 207 and 152 alleles were detected in *P. euphratica* and *P. pruinosa*, respectively. Following our analyses based on these microsatellite loci, 54 distinct MLGs were detected from the 122 samples of *P. pruinosa*, and 203 MLGs were detected from 551 samples of *P. euphratica*, with the clonal diversity (R) ranging from 0.077–0.821 within populations of *P. pruinosa*, and 0–0.947 within populations of *P. euphratica* (Table 1 and Additional file 1: Figure S1).

After removing replicates belonging to identical MLGs, the overall gene diversity ( $H_T$ ) was similar for *P. euphratica* (0.594) and *P. pruinosa* (0.565). The mean values of  $H_O$  and  $H_E$  respectively across populations

were 0.538 and 0.558 for *P. euphratica*, and 0.484 and 0.560 for *P. pruinosa*. After using a sequential Bonferroni correction, only one *P. euphratica* population that coexisted with *P. pruinosa* individuals, ALe, had a fixation index ( $F_{IS}$ ) that significantly deviated from zero, suggesting that all populations except ALe are under a Hardy–Weinberg equilibrium. No  $F_{IS}$  significantly deviated from zero for each population of *P. pruinosa*; but there was a significant deviation for the species as a whole, possibly reflecting the existence of a Wahlund effect (Table 1).

When all *P. euphratica* populations were divided into three regions, the regional genetic diversity, as  $A_R$ ,  $P_A$ ,  $H_O$ , and  $H_S$ , for northern Xinjiang was relatively high compared with southern Xinjiang and the QGM region; while the  $F_{IS}$  value was slightly higher for southern Xinjiang than the other two regions (Table 2).

### Population differentiation and structure

The interspecific genetic differentiation based on the Weir and Cockerham [54] estimator was moderate and significant across all microsatellite loci ( $\theta = 0.328 \pm 0.073$ ,  $p < 0.01$ ). Differentiation among populations was low within both species, but was relatively higher in *P. euphratica* ( $\theta = 0.062 \pm 0.005$ ,  $p < 0.01$ ) than in *P. pruinosa* ( $\theta = 0.009 \pm 0.007$ ,  $p > 0.05$ ). For *P. euphratica*, the  $F_{ST}$  values among populations within northern Xinjiang (0.047) were higher than those within both southern Xinjiang (0.020) and the QGM region (0.028, Table 2).

The STRUCTURE output for all MLGs showed that both likelihood ( $\ln P(D)$ ) and  $\Delta K$  supported the existence of two clusters (See Additional file 1: Figure S4a and b), which corresponded well with our morphological assignment to *P. euphratica* and *P. pruinosa* (Fig. 1a). The *P. euphratica* populations could be further classified into two clusters (See Additional file 1: Figure S4c and d), which corresponded approximately to populations from southern and northern Xinjiang, with the population BEJa from northern Xinjiang being an exception (Fig. 1b). Almost all populations from the QGM region were a mixture between the two Xinjiang clusters.

### Population history and gene flow

A comparison of the posterior probabilities (PPs) of the five scenarios using local linear regression indicated that Scenario 5 was the most likely (PP = 0.97, 95% CIs: 0.95–0.98; See Additional file 1: Figure S3). For Scenario 5, the median values of  $t_1$  and  $t_2$  were 336 generations (95% CIs: 65–1610) and 18,500 generations (95% CIs: 8100–32,800), respectively (Table 3). The ra composition of the southern Xinjiang group was 34.2%, while the median values of the effective population sizes of northern Xinjiang, southern Xinjiang, the QGM region, and ancestral population were 84,000, 80,400, 17,700, and 2780, respectively.

**Table 1** Genetic diversity of each population of *Populus pruinosa* and *P. euphratica*

Code	Nuclear Microsatellites							Chloroplast sequence		
	$N_s$	MLGs	R	A	$H_o$	$H_E$	$F_{IS}$	$N_c$	$N_h$	Hap_div
<i>P. pruinosa</i>	122	54	0.438	152	0.484	0.560	0.112*	58	8	0.730
MYp	15	8	0.500	67	0.493	0.590	0.175	9	3	0.667
BCp	14	2	0.077	35	0.441	0.480	0.118	4	2	0.500
Alp	16	8	0.467	68	0.477	0.520	0.089	12	3	0.530
GMp	40	33	0.821	135	0.539	0.595	0.094	26	5	0.668
Ylp	17	3	0.125	46	0.471	0.549	0.172	7	2	0.476
<i>P. euphratica</i>	551	203	0.368	207	0.538	0.558	0.046	225	20	0.705
MYe	16	8	0.467	77	0.596	0.563	-0.062	10	1	0.000
BCe	14	8	0.538	79	0.581	0.588	0.013	11	4	0.691
Ale	17	8	0.438	72	0.412	0.541	0.251 <sup>a</sup>	9	3	0.417
MF	16	10	0.600	74	0.505	0.539	0.065	11	3	0.473
PSH	16	3	0.133	49	0.569	0.541	-0.064	9	2	0.556
QM	18	1	0.000	24	-	-	-	2	1	0.000
YPH	20	9	0.421	73	0.473	0.504	0.065	9	2	0.222
RQ	16	10	0.600	70	0.444	0.512	0.140	11	1	0.000
ShY	17	10	0.588	99	0.547	0.585	0.068	10	2	0.200
KEL	14	12	0.846	65	0.477	0.492	0.032	10	3	0.378
WL	25	13	0.500	75	0.511	0.511	0.000	6	1	0.000
LT	7	1	0.000	26	-	-	-	1	1	-
YLe	13	6	0.417	69	0.549	0.579	0.057	7	4	0.810
ChJ	16	10	0.600	76	0.527	0.576	0.088	11	4	0.491
KLM	21	4	0.150	62	0.559	0.599	0.077	5	3	0.800
MGCa	15	2	0.071	51	0.549	0.612	0.125	4	2	0.500
MGCb	17	1	0.000							
BEJa	20	11	0.526	84	0.512	0.553	0.077	11	4	0.600
BEJb	20	7	0.316	58	0.626	0.575	-0.096	2	2	1.000
HBH	16	4	0.200	52	0.618	0.545	-0.161	4	2	0.500
BLK	3	1	0.000	70	0.588	0.593	0.099			
BLST	5	1	0.000					12	4	0.652
BLBQ	9	4	0.735							
QHX	3	1	0.000	25	-	-	-	3	1	0.000
MLX	20	1	0.000	24	-	-	-	12	1	0.000
YWX	24	15	0.625	96	0.539	0.582	0.099	12	4	0.803
TLH1	16	1	0.000	41	0.559	0.598	0.095	8	4	0.750
TLH2	35	1	0.000							
AKS	11	5	0.400	54	0.612	0.527	-0.185	5	2	0.400
JQ	23	1	0.000	23	-	-	-	2	1	0.000
DH	28	1	0.000	24	-	-	-	7	1	0.000
JT	20	19	0.947	83	0.489	0.509	0.040	7	3	0.762
EQN	20	14	0.684	94	0.538	0.584	0.082	14	4	0.648
Total	673	257	-	228	-	-	-	268	25	-

Number of samples genotyped ( $N_s$ ), number of multilocus genotypes (MLGs) identified, clonal diversity (R) and total number of alleles on the entire sample (A), expected heterozygosity ( $H_e$ ) and deviation from Hardy–Weinberg equilibrium ( $F_{IS}$ , <sup>a</sup>significant after a 1000 permutation test) with the 17 nuclear microsatellites. Number of samples sequenced ( $N_c$ ), number of haplotypes ( $N_h$ ) and haplotype diversity (Hap\_div) with the chloroplast *trnK* sequences

**Table 2** Comparison of genetic diversity and differentiation among three geographical populations of *Populus euphratica*

Group	Microsatellite						Chloroplast			
	$A_R$	$P_A$	$H_O$	$H_S$	$F_{IS}$	$F_{ST}$	$N_H$	$H_T$	$H_S$	$G_{ST}$
Northern Xinjiang	7.25	1.39	0.561	0.583	0.038	0.047	13	0.828	0.594	0.283
Southern Xinjiang	6.76	0.98	0.506	0.536	0.057	0.020	9	0.344	0.312	0.093
QGM region	6.98	0.71	0.525	0.542	0.032	0.028	7	0.646	0.423	0.345

$A_R$ , Allelic richness;  $P_A$ , private allelic richness;  $H_O$ , observed heterozygosity;  $H_S$ , gene diversity;  $F_{IS}$ , fixation index;  $F_{ST}$ , among population differentiation. The analyses of  $A_R$  and  $P_A$  are based on 50 gene copies

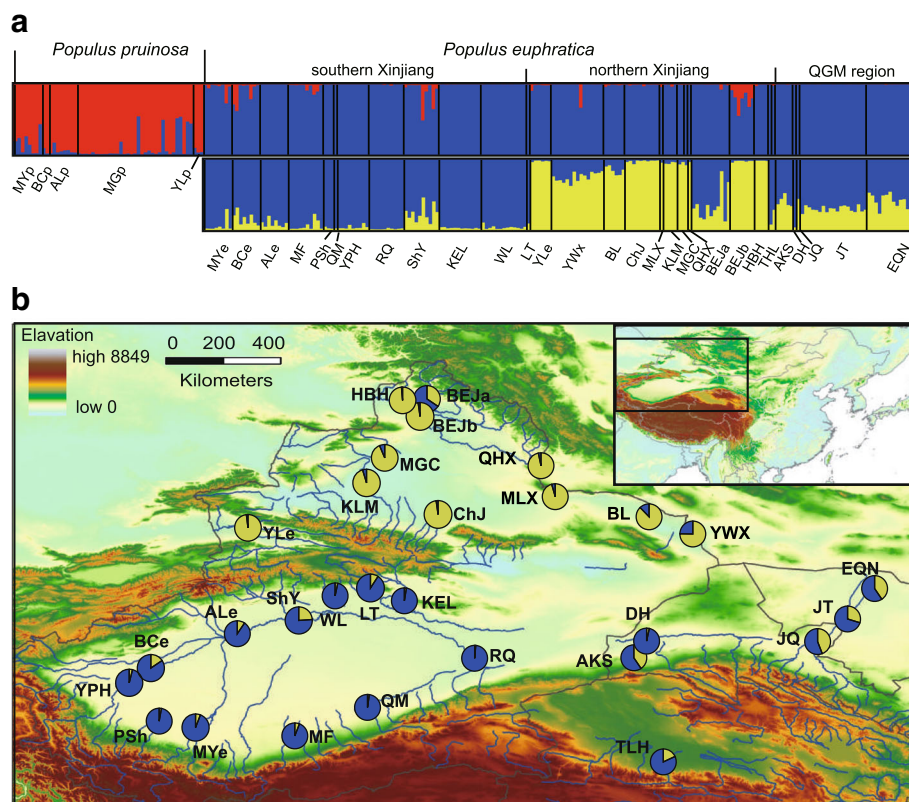
Estimates of gene flow generated using Migrate were low between northern and southern Xinjiang,  $4Nm_{north \rightarrow south} = 0.91$  and  $4Nm_{south \rightarrow north} = 1.39$ . The highest level of migration occurred from northern Xinjiang to the QGM region ( $4Nm_{north \rightarrow QGM} = 6.60$ , 95% CIs: 6.34-6.86), followed by the QGM region to southern Xinjiang ( $4Nm_{QGM \rightarrow south} = 4.98$ , 95% CIs: 4.72-5.25), resulting in a relatively higher gene flow from northern Xinjiang to southern Xinjiang via the QGM region than the reverse ( $4Nm_{south \rightarrow QGM} = 3.03$ , 95% CIs: 2.86-3.20 and  $4Nm_{QGM \rightarrow north} = 4.46$ , 95% CIs: 4.28-4.64; Fig. 2a).

The BayesAss analysis showed that the mean contemporary gene flow (m) among pairs of the three groups

ranged from 0.0042 to 0.3137 (Fig. 2b). The highest level of migration occurred from the QGM region to southern Xinjiang ( $m_{QGM \rightarrow south} = 0.3137 \pm 0.0136$ ), followed by migration from northern Xinjiang to the QGM region ( $m_{north \rightarrow QGM} = 0.3044 \pm 0.0124$ ). Other pairs of gene flow were all lower than 0.0066 (Fig. 2b), also suggesting a higher contemporary gene flow from northern to southern Xinjiang via the QGM region rather than the reverse.

**Chloroplast DNA variation and phylogenetics**

The aligned cpDNA *trnK* data matrix was 2528 bp in length. A total of 25 haplotypes were identified in the 38



**Fig. 1** Bayesian estimation of the proportion of genetic clusters for each multilocus genotype (MLG) and population using STRUCTURE software. Analyses were conducted based on 17 microsatellite loci. (a) Proportion of genetic clusters at  $K = 2$  for 257 MLGs of *Populus euphratica* and *P. pruinosa*, and 203 MLGs of *P. euphratica*. The smallest vertical bar represents one MLG. The assignment proportion of each MLG into the population clusters is shown along the y-axis. (b) Geographical distribution of the two genetic clusters and composition of the genetic cluster in each population of *P. euphratica*. The map was created using the ArcMap package in ArcGIS ver. 9.2 (<http://www.esri.com/software/arcgis>)

**Table 3** Demographic approximate Bayesian computation models for *Populus euphratica* at scenario 5

Parameter	median	5%	95%
N1	8.40E + 04	5.53E + 04	9.82E + 04
N2	8.04E + 04	5.08E + 04	9.73E + 04
N3	1.77E + 04	4.39E + 03	6.77E + 04
Na	2.78E + 03	4.99E + 02	1.15E + 04
t1	3.36E + 02	6.50E + 01	1.61E + 03
t2	1.85E + 04	8.10E + 03	3.28E + 04
ra	3.42E-01	7.22E-02	7.75E-01
B <sub>pmic_1</sub>	5.26E-05	3.20E-05	8.39E-05
pmic_1	6.58E-01	4.37E-01	8.71E-01

N1, N2, and N3, current effective population size of gene pools from the northern Xinjiang, southern Xinjiang, and the QGM region, respectively; Na, ancestral effective population size; t1–t2, estimated times of the different events depicted in Additional file 1: Figure S3 (in generations); ra, admixture rate of gene pools from southern Xinjiang; B<sub>pmic\_1</sub>, estimated microsatellite mutation rate; pmic\_1, the parameter of the geometric distribution

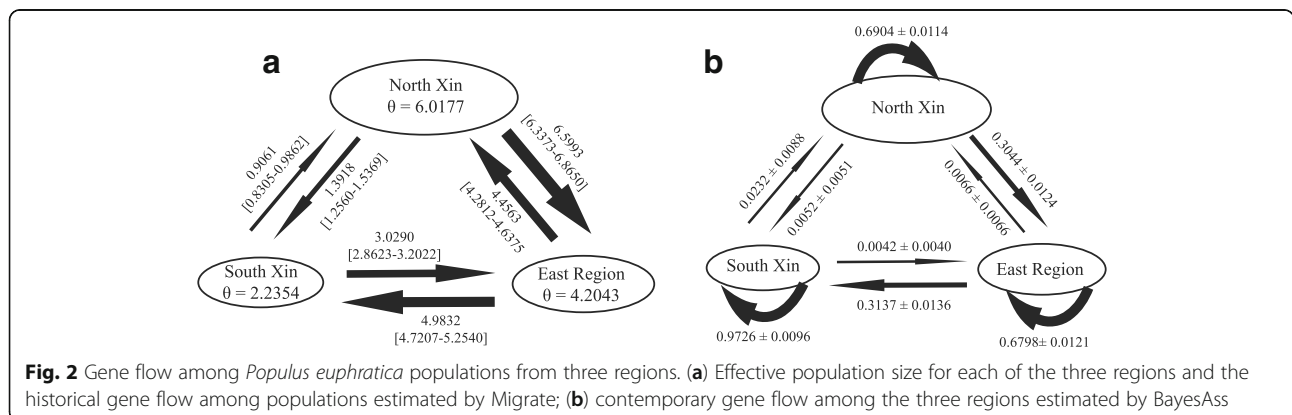
*P. euphratica* and *P. pruinosa* populations, based on 18 nucleotide substitutions and six indels (See Additional file 1: Table S4). Twenty haplotypes (H01–H20) occurred in *P. euphratica*, eight (H15, H19–H25) occurred in *P. pruinosa*, and three were shared by the two species (H15, H19, and H20). Another haplotype that differed from all 25 haplotypes mentioned above (by at least nine substitutions) was identified for the three *P. ilicifolia* individuals.

Haplotype diversity ( $H_d$ ) ranged from 0.476–0.668 and 0–0.810 within the five *P. pruinosa* populations, and 33 *P. euphratica* populations, respectively (Table 1). The level of total genetic diversity,  $H_T$ , was 0.730 and 0.705, across populations of *P. pruinosa* and *P. euphratica*, respectively. The population differentiation was larger for *P. euphratica* ( $G_{ST} = 0.391$ ) than it was for *P. pruinosa* ( $G_{ST} = 0.183$ ). Among the three regions of *P. euphratica*, the population from southern Xinjiang had the lowest chloroplast diversity and population divergence (Table 2).

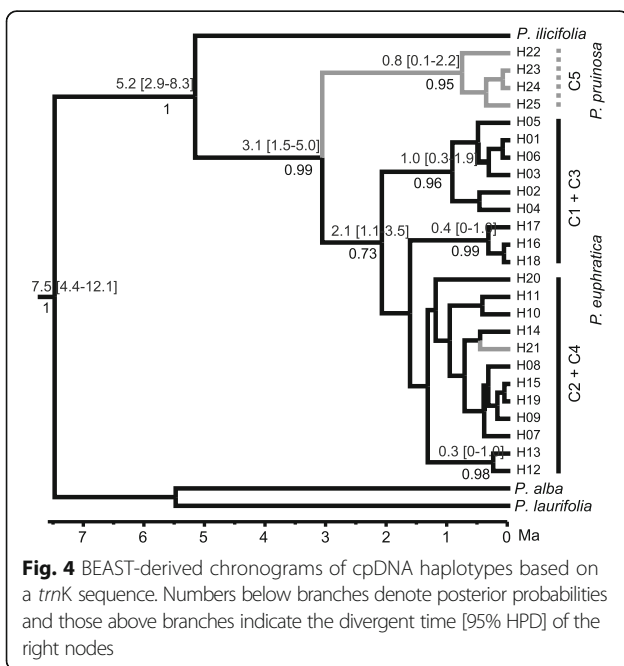
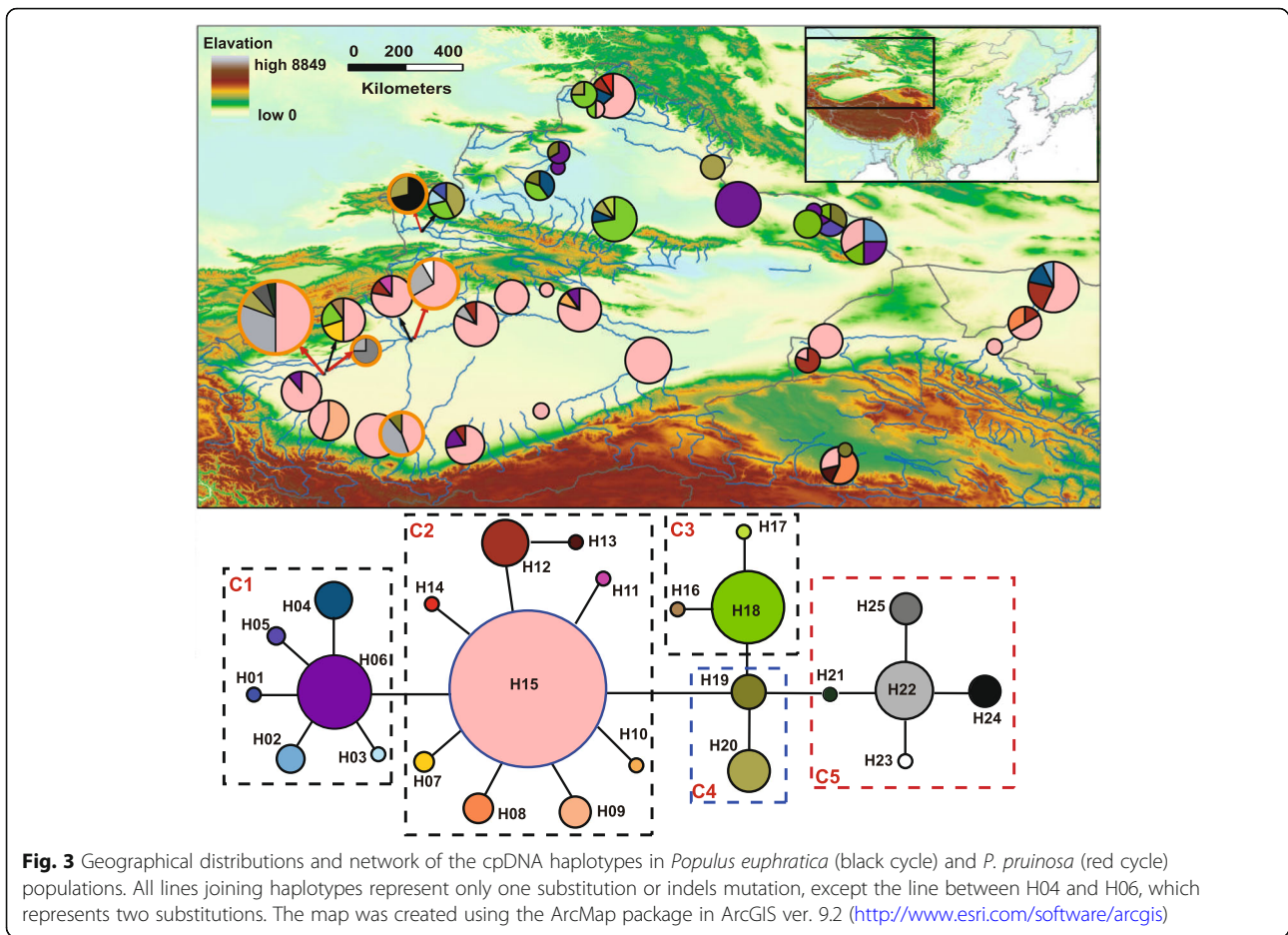
All 25 of the cpDNA haplotypes from *P. euphratica* and *P. pruinosa* were connected into a network by one

mutation between each other (except two between H04 and H06; Fig. 3), and the haplotype from *P. ilicifolia* was connected with H15 by 11 mutations (not shown). The network could be roughly classified into five clades. Clade 1 (H01–H06), Clade 2 (H07–H15), and Clade 3 (H16–H18) were composed of haplotypes that only occurred in *P. euphratica*, except H15, which was located in the center of the network and was shared by the two species and was widespread (50.5%). Clade 4 (H19–H20) was located in the center of the network and was also shared by the two species, and Clade 5 (H21–H25) was composed of haplotypes that only occurred in *P. pruinosa*. The haplotype distribution in *P. euphratica* populations from northern and southern Xinjiang was clearly different. Haplotypes from Clade 1 and Clade 3 were mainly distributed in the northern Xinjiang populations, while haplotypes from Clade 2 were mainly found in southern Xinjiang populations and eastern populations. H15 that mainly occurred in southern Xinjiang and QGM regions was found in three northern Xinjiang populations: YWX, BEJa, and BEJb.

The phylogenetic tree showed that, when *P. alba* and *P. laurifolia* were used as outgroups, all haplotypes from *P. euphratica*, *P. pruinosa*, and *P. ilicifolia* formed a monophyletic group with a support probability of 1 (Fig. 4). Within the monophyletic group, the haplotype identified in *P. ilicifolia* was divergent from all other haplotypes, with a support probability of 1, at around 5.2 Ma (95% CIs: 2.9–8.3 Ma). All other haplotypes could be roughly classified into four lineages, with a support probability of higher than 0.7. The four *P. pruinosa* private haplotypes (H22–H25), that belonged to Clade 5 of the haplotype network, formed a lineage that was divergent at around 3.1 Ma (95% CIs: 1.5–5.0 Ma). Haplotypes from northern Xinjiang formed two lineages, which corresponded well to Clade 1 and Clade 3 in the haplotype network, and were divergent at around 2.1 Ma (95% CIs: 1.0–3.5 Ma) and at around 1.5–2.0 Ma, respectively. All other haplotypes formed the fourth lineage.

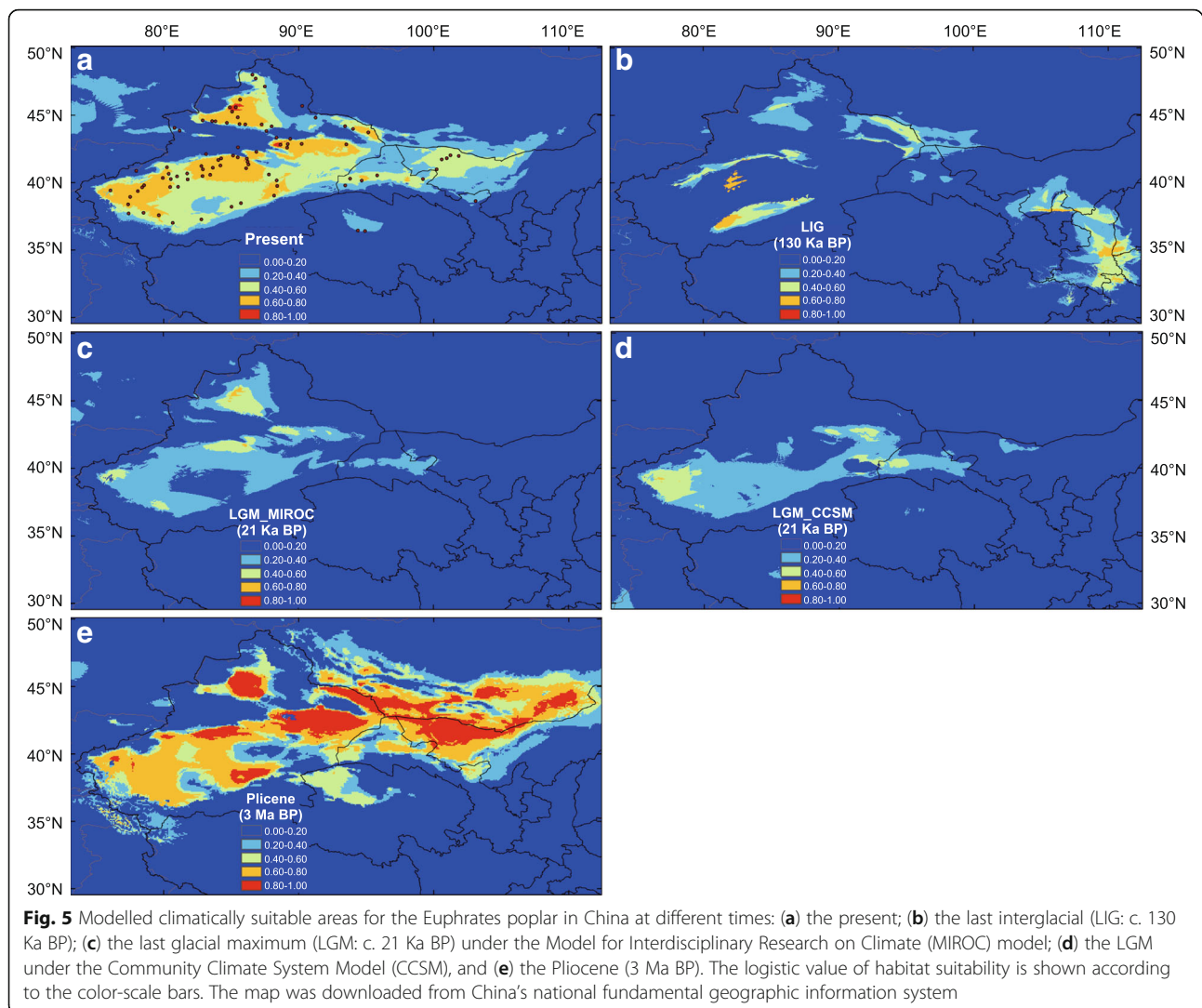






### Present and past distribution of desert poplar

The ENM analysis based on all records of *P. euphratica* predicted completely separate potential distributions in China and those regions west of China, e.g., central Asia, Europe, and northern Africa, in all models, except for a slight connection at the western edge of the QTP (See Additional file 1: Figure S5). To specify the suitability of the estimation of the Euphrates poplar in China, we considered the ENM results based on Chinese records only. This resulted in a high predictive power, with an area under the curve of  $0.941 \pm 0.054$ . The prediction of a model over the present bioclimatic conditions showed a good to moderate suitability of the species' extant distribution in Xinjiang province, northern Gansu province, eastern Inner Mongolia province, and a few low-suitability distributions in Qinghai province (Fig. 5a), which is consistent with the present distribution of the Euphrates poplar in China. Both the MIROC model (Fig. 5b) and the CCSM model (Fig. 5c) predicted only a few moderate suitability distributions scattered in Xinjiang province at the LGM. The prediction of the LIG also showed some moderate and scattered suitability distributions, and suitable habitats still occurred in



regions further to the east (Fig. 5d). Furthermore, highly suitable habitats were predicted to be widely distributed in northwest China under the bioclimatic conditions of 3 Ma (Fig. 5e), indicating a much wider distribution range of the Euphrates poplar during the Pliocene. The ENM analysis for *P. pruinosa* also had a high predictive power, with an area under the curve of  $0.966 \pm 0.066$ . The predicted distribution range was also much wider during the Pliocene than at the present, and was slightly contracted during the LGM, but more restricted during the LIG (See Additional file 1: Figure S6).

## Discussion

### The Tianshan Mountains as a geological barrier to population divergence

Both the cpDNA *trnK* sequence and microsatellite variations demonstrate the apparent differentiation between northern and southern Xinjiang populations of Euphrates poplar, suggesting a long period of isolation between

these populations. Our analysis based on nuclear microsatellites found very low levels historic and concurrent gene flow between Euphrates poplar populations from these two regions. Moreover, the limited gene flow between northern and southern Xinjiang probably occurred via the QGM region, and a higher gene flow was found from northern to southern Xinjiang via QGM than in the reverse direction (Fig. 2). Clearly, the formation of the Tianshan Mountains limited both seed dispersal and pollen flow among populations from these two regions.

Chronostratigraphic framework results from various locations within the Tianshan Mountains generally show that deformation occurred during four time intervals: early Miocene (25–20 Ma), Middle Miocene (17–15 Ma), early late Miocene (11–10 Ma), and late Miocene (7–5 Ma) [7]. Our results suggest that the chloroplast lineages distributed in northern Xinjiang became divergent from those in southern Xinjiang during

the early Pleistocene, ca. 1.5–2.1 Ma (Fig. 4). Based on microsatellites and using DIYABC software, it was estimated that the divergence time between the northern and southern Xinjiang groups ( $t_2$ ) was 18,500 generations (95% CIs: 8100–32,800) (Table 3), which was converted to 0.37 Ma (95% CI: 0.16–0.65 Ma), assuming 20 years' generation time. These estimated divergence times were much later than the latest Miocene deformation of the Tianshan Mountains suggested by the chronostratigraphic framework. One possible explanation for this is that the latest deformation of the Tianshan Mountains was much later than previously suggested, for example during the Pleistocene rather than the Miocene [55, 56]. However, Renner (2016) reported that the majority of phylogeny-cum-biogeography studies tended to support recent QTP uplift phases (e.g., 20–0.5 Ma). The young species divergences inferred in these studies may not be simply driven by the QTP uplift, but may reflect recent re-colonization of the plateau from possible refugia or be due to local habitat differences [57]. Thus, the later estimated divergence time in our current analysis suggested that the gene flow between Euphrates poplar populations from northern and southern Xinjiang regions had not been fully prevented by the latest Miocene deformation of the Tianshan Mountains. When Euphrates poplar was distributed widely during the Pliocene (Fig. 5e), gene flow between these two regions could still occur via the QGM region (Fig. 2).

#### Influence of quaternary climatic oscillations

In northwest China, despite the absence of major Quaternary glaciations, significant climatic oscillations still occurred. During the Pleistocene glaciations, species at lower latitudes were subjected to extreme aridity as well as lower temperatures (Willis and Niklas, 2004). Sandy desert and gobi (stony desert) expansion has caused the habitat fragmentation of desert plants populations [6]. We suggest that the historical climate has greatly influenced the population demography of the Euphrates poplar in northwest China. First, the ENM estimated that suitable habitats of the Euphrates poplar in northwest China was much wider during the Pliocene (Fig. 5e), but decreased substantially during the LGM (Fig. 5b and c) and the LIG (Fig. 5d). This suggested that both the glacial (driven by EAWM) and interglacial (driven by the EASM) climates could have greatly contracted the habitat of this species and finally fragmented its distribution range (Fig. 5b, c, and d). Second, the 'star-like' phylogeny of haplotypes also supports a population bottleneck in the historical demography of *P. euphratica*, with the occurrence of a single, common, and often frequent ancestral haplotype (H15) in all populations of southern Xinjiang suggesting a

significant historical bottleneck. Finally, the DIYABC analysis suggested that current Euphrates poplar distributions in the QGM region were created by an admixture of the northern and southern Xinjiang gene pools 336 generations ago (95% CIs: 65–1610; Scenario 5 see Additional file 1: Figure S3). Using 20 years as the generation time, an admixture of the two gene pools dates back to 6720 years ago, corresponding to the postglacial period. Thus, the current Euphrates poplar distribution in the QGM region was because of postglacial population expansion from the northern and southern Xinjiang populations. Because the gene flow between northern and southern Xinjiang mainly occurred via the QGM region (Fig. 2), the absence of the distribution in the QGM region during the glacial period would have further blocked gene flow between the populations of these two regions. Therefore, we conclude that when the deformation of the Tianshan Mountains limited the gene flow of *P. euphratica* populations from northern and southern Xinjiang, the Pleistocene climatic oscillations further accelerated the population divergence between these two regions. The cycles of the monsoonal climatic oscillations likely played a key role in the habitat fragmentation and intraspecific divergence of *P. euphratica*. The higher gene flow from northern to southern Xinjiang via the QGM than in the reverse direction (Fig. 2) indicated that the EAWM accelerated the gene flow of the desert poplar along the latitudinal direction. Similar results were found in another desert plant, *Reaumuria soongarica* [58].

The higher genetic diversity in northern Xinjiang than in southern Xinjiang in both the nuclear and chloroplast genome of *P. euphratica* (Table 2) suggested a smaller influence of historical climatic oscillations in the former region, although artificial introductions might also have contributed, e.g., population BEJa, whose chloroplast haplotypes and ancestry both belonged to southern Xinjiang according to results of Bayesian clustering analysis. The high differentiation among populations from northern Xinjiang suggested multiple historical isolations in the distribution of *P. euphratica*. The heterogeneous geology in northern Xinjiang might have facilitated the existence of multiple glacial refugia. The Ili Valley, located near the juncture between the northern and southern branches of the Tianshan Mountains, has been shown to be a glacial refugium for plants [59]. In our analysis, *P. euphratica* in Ili Valley also held distinct chloroplast haplotypes (Fig. 2), supporting the existence of a glacial distribution. The Altay-Tianshan Mountains that were found to include glacial refugia for many plants [6] may also be helpful for the glacial distribution of *P. euphratica*. Climatic oscillations have caused extreme aridity and the expansion of sandy deserts in southern Xinjiang regions [6], which would have further led to historical habitat shrinkage for *P. euphratica*.

### Genetic diversity of desert poplar and conservation applications in China

Consistent with a previous study [18], our analyses found that *P. euphratica* differed from *P. pruinosa* in both the nuclear and chloroplast genome, although there were several hybrids and a shared ancestry of cpDNA haplotypes (Figs. 1 and 3). For both species, clone diversity was low within many populations, suggesting low regeneration by seeds/seedlings of these two desert species in many locations. This was particularly true of the DH population that is located in the Euphrates poplar forest in the Dunhuang reserve, and for TLHa and TLHb, the only two high-altitude populations located in the Chaidamu Basin, with only one genet found in each of them. However, for these populations, a larger sample size and greater distance between individuals were represented during sampling. Both *P. euphratica* and *P. pruinosa* can regenerate from seed/seedlings and root suckers [20, 60], while seedlings can emerge only after flooding events [61]. Cao [60] suggested that water shortages and river channeling due to water usage and altered river flows might have resulted in no safe sites on river banks for seed germination in the National Natural Reserve of *P. euphratica*, in Ejina. This would lead to a failure of *P. euphratica* to regenerate from seed, with root suckers being the main source of recruitment in some fields [60]. Thus, the low clonal diversity might reflect a water shortage and low groundwater table in these regions. To preserve and restore desert poplars, it is necessary to produce wet conditions favoring seed germination and establishment by regulating stream flow, for example, releasing sufficient water during peak seed rain [60].

We found that the overall gene diversity ( $H_T$ ) based on MLGs was similar for *P. euphratica* (0.594) and *P. pruinosa* (0.565). The genetic diversity at a population level was also comparable between the two species ( $H_O = 0.484$  and  $H_E = 0.560$  for *P. euphratica* and  $H_O = 0.538$  and  $H_E = 0.558$  for *P. pruinosa*). These values were lower than those stated in a previous study of the two species [62], possibly due to the difference in polymorphic loci selection used in the analyses. However, our genetic diversity estimates of these two desert poplars are still at an intermediate level compared with other congeneric species that have been estimated using microsatellite markers, such as *P. alba* ( $H_O = 0.341$ ,  $H_E = 0.368$ ) and *P. tremula* ( $H_O = 0.483$ ,  $H_E = 0.492$ ) [63], and *P. nigra* ( $H_O = 0.70$ ,  $H_E = 0.73$ ) [64]. This suggests that genetic diversity for the two endangered species in northwest China is not as low as we originally assumed.

Among *P. euphratica* from the three regions, populations from northern Xinjiang province held the highest nuclear and chloroplast genetic diversity, while populations from southern Xinjiang held the lowest genetic diversity (Table 2).

The higher genetic diversity in northern Xinjiang reflects the larger effective population size and greater endemic gene resources there than in southern Xinjiang, although the latter region represents the widest Euphrates poplar distribution globally [16]. Furthermore, in northwest China, natural desert poplar reserves were mainly established in the southern Xinjiang region, e.g., Tarim national natural reserve, and in the QGM region, e.g., Ejina national natural reserve and the Dunhuang reserve. None were established in the northern Xinjiang region. Thus, we suggest that to restore the genetic resources of the Euphrates poplar, more effort is needed to protect the populations in the northern Xinjiang region.

### Conclusions

Studies in the Hengduan Mountains and adjacent regions in southwest China found that both orogeny and past climatic changes have contributed to high plant diversity [2–4]. However the roles of geology and climatic oscillations in driving the population demography and divergence of plants on the northern edge of the QTP in arid northwest China were still unknown. Our current study found that the orogeny of the Tianshan Mountains has somewhat impeded gene flow between *P. euphratica* populations from northern and southern Xinjiang; while Quaternary climatic aridification has caused significant habitat fragmentation and population contraction, which further accelerated the population divergence. These results provide a new insight into the low diversity of plants in this arid region. In addition, we also found that a water shortage and low groundwater table have resulted in a low regeneration rate of seed/seedlings in many populations of both *P. euphratica* and *P. pruinosa*. To better restore the genetic resources of the desert poplars, more effort is needed to encourage seed germination and seedling establishment, and to protect populations distributed in the northern Xinjiang region, which hold the highest genetic diversity and a large amount of endemic gene diversity.

### Additional file

**Additional file 1: Figure S1.** Sampling location of *Populus euphratica* and *P. pruinosa* populations. **Figure S2.** Frequency distribution of the pairwise distance distribution between individuals based on multilocus genotypes. **Figure S3.** The five scenarios tested in the DIYabc analysis. **Figure S4.** Inference of the most probable number of clusters (K) using STRUCTURE software. **Figure S5.** Modelled climatically suitable areas for Euphrates poplar. **Figure S6.** Modelled climatically suitable areas for *P. pruinosa*. **Table S1.** Description of *Populus pruinosa* and *P. euphratica* populations analysed. **Table S2.** Description and references of the 17 microsatellite loci analysed for this study. **Table S3.** Diversity and differentiation for the 17 microsatellite loci analysed in *Populus euphratica* and *P. pruinosa*. **Table S4.** Variable sites of the aligned sequences of chloroplast DNA fragments in 25 haplotypes of *Populus euphratica* and *P. pruinosa* in northwest China. (DOCK 2025 kb)

## Abbreviations

CCSM: Community Climate System Model; CIs: Confidence intervals; cpDNA: Chloroplast DNA; EASM: the East Asian summer monsoons; EAWM: the East Asian winter monsoons; ENM: Ecological Niche Modeling; LGM: Last glacial maximum; LIG: Last interglacial; MIROC: the Model for Interdisciplinary Research on Climate; ML: Maximum likelihood; MLGs: Multilocus genotypes; PCR: Polymerase chain reaction; PP: Posterior probabilities; QGM: Qinghai province, Gansu province, and Inner Mongolia province; QTP: Qinghai-Tibetan Plateau

## Acknowledgements

We are grateful to our many colleagues in forestry research from Xinjiang province for their help during sample collections.

## Funding

This project was supported by the Fundamental Research Funds for the Central Non-profit Research Institution of Chinese Academy of Forestry (RIF2008-02) to YFZ; WTW was supported by the National Natural Science Foundation of China (31560127); the publication of this article was in part funded by the National Natural Science Foundation of China (31670666).

## Availability of data and materials

All sequence data were deposited in GenBank under accession numbers KY002202–KY002230.

Microsatellite data matrix was deposited at Dryad: doi:<https://doi.org/10.5061/dryad.v95q1>.

## Authors' contributions

YFZ and JGZ wrote the manuscript. YFZ performed laboratory work and data analyses. JGZ developed the research idea with the support of the other authors and obtained funding for the research. WTW performed the ENM simulation. YFZ, JGZ, AB and ZQJ sampled the populations. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>State Key Laboratory of Tree Genetics and Breeding, Research Institute of Forestry, Chinese Academy of Forestry, NO. 1 Dongxiaofu, Xiangshan road, Haidian district, Beijing 100091, China. <sup>2</sup>Collaborative Innovation Center of Sustainable Forestry in Southern China, Nanjing Forestry University, No. 159 Longpan road, Nanjing 210037, China. <sup>3</sup>Key Laboratory of Tree Breeding and Cultivation of the State Forestry Administration, Research Institute of Forestry, Chinese Academy of Forestry, No. 1 Dongxiaofu, Xiangshan road, Haidian district, Beijing 100091, China. <sup>4</sup>Forest Research Institute of the Altai Region, No. 93 Jiefang South Road, Altai, Xinjiang 836500, China. <sup>5</sup>School of Mathematics and Computer Science, Northwest University for Nationalities, No. 1 Northwest new village, Lanzhou 730030, Gansu, China. <sup>6</sup>Institute of Desertification Studies, Chinese Academy of Forestry, Xiangshan road, Haidian district, Beijing 100091, China.

Received: 30 August 2017 Accepted: 17 May 2018

Published online: 25 May 2018

## References

- Liu JQ, Sun YS, Ge XJ, Gao LM, Qiu YX. Phylogeographic studies of plants in China: advances in the past and directions in the future. *J Syst Evol.* 2012;50:267–75.
- Qiu YX, Fu CX, Comes HP. Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of quaternary climate and environmental change in the world's most diverse temperate flora. *Mol Phylogenet Evol.* 2011;59:225–44.
- Favre A, Päckert M, Pauls SU, Jähnig SC, Uhl D, Michalak I, Muellner-Riehl AN. The role of the uplift of the Qinghai-Tibetan plateau for the evolution of Tibetan biotas. *Biol Rev.* 2015;90:236–53.
- Xing Y, Ree RH. Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. *Proc Natl Acad Sci U S A.* 2017;114:E3444–51.
- Meng HH, Su T, Gao XY, Li J, Jiang XL, Sun H, Zhou ZK. Warm–cold colonization: response of oaks to uplift of the Himalaya–Hengduan Mountains. *Mol Ecol.* 2017;26:3276–3294.
- Meng H-H, Gao X-Y, Huang J-F, Zhang M-L. Plant phylogeography in arid Northwest China: retrospectives and perspectives. *J Syst Evol.* 2015;53:33–46.
- Tang Z, Yang S, Qiao Q, Yin F, Huang B, Ding Z. A high-resolution geochemical record from the Kuche depression: constraints on early Miocene uplift of south tian Shan. *Palaeogeogr Palaeoclimatol Palaeoecol.* 2016;446:1–10.
- Miao Y, Herrmann M, Wu F, Yan X, Yang S. What controlled mid–late Miocene long-term aridification in Central Asia?—global cooling or Tibetan plateau uplift: a review. *Earth-Sci Rev.* 2012;112:155–72.
- Lu H, Guo Z. Evolution of the monsoon and dry climate in East Asia during late Cenozoic: a review. *Sci China Earth Sci.* 2014;57:70–9.
- An Z, Kutzbach JE, Prell WL, Porter SC. Evolution of Asian monsoons and phased uplift of the Himalaya–Tibetan plateau since late Miocene times. *Nature.* 2001;411:62–6.
- Ding ZL, Derbyshire E, Yang SL, Sun JM, Liu TS. Stepwise expansion of desert environment across northern China in the past 3.5 ma and implications for monsoon evolution. *Earth Planet Sci Lett.* 2005;237:45–55.
- Yang X, Scuderi LA. Hydrological and climatic changes in deserts of China since the late Pleistocene. *Quat Res.* 2010;73:1–9.
- Dang R, Pan X, Gu X. Floristic analysis of spermatophyte genera in the arid deserts area in north-West China. *Guangxi Zhiwu.* 2002;22:121–8.
- Jiang X-L, Zhang M-L, Zhang H-X, Sanderson SC. Phylogeographic patterns of the *Aconitum nemorum* species group (Ranunculaceae) shaped by geological and climatic events in the Tianshan Mountains and their surroundings. *Plant Syst Evol.* 2014;300:51–61.
- Xu Z, Zhang M-L. Phylogeography of the arid shrub *Atraphaxis frutescens* (Polygonaceae) in northwestern China: evidence from cpDNA sequences. *J Hered.* 2015;106:184–95.
- Wang SJ, Chen BH, Li HQ. *Euphrates Poplar Forest*. Beijing: China Environmental Science Press; 1995.
- Wang J, Wu Y, Ren G, Guo Q, Liu J, Lascoux M. Genetic differentiation and delimitation between ecologically diverged *Populus euphratica* and *P. pruinosa*. *PLoS One.* 2011;6:e26530.
- Wang J, Källman T, Liu J, Guo Q, Wu Y, Lin K, Lascoux M. Speciation of two desert poplar species triggered by Pleistocene climatic oscillations. *Heredity.* 2014;112:156–64.
- Bruelheide H, Jandt U, Gries D, Thomas FM, Foetzki A, Buerkert A, Gang W, Ximing Z, Runge M. Vegetation changes in a river oasis on the southern rim of the Taklamakan Desert in China between 1956 and 2000. *Phytocoenologia.* 2003;33:801–18.
- Hukin D, Cochard H, Dreyer E, Le Thiec D, Bogaert-Triboulet MB. Cavitation vulnerability in roots and shoots: does *Populus euphratica* Oliv., a poplar from arid areas of Central Asia, differ from other poplar species? *J Exp Bot.* 2005;56:2003–10.
- Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science.* 2006;313:1596–604.
- Wu YX, Wang J, Liu JQ. Development and characterization of microsatellite markers in *Populus euphratica* (Populaceae). *Mol Ecol Resour.* 2008;8:1142–4.
- Schuelke M. An economic method for the fluorescent labeling of PCR fragments. *Nat Biotechnol.* 2000;18:233–4.
- Demesure B, Sodzi N, Pétit RJ. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Mol Ecol.* 1995;4:129–34.
- Zeng YF, Zhang JG, Duan AG, Abuduhaiti B. Genetic structure of *Populus* hybrid zone along the Irtysh River provides insight into plastid-nuclear incompatibility. *Sci Rep.* 2016;6:28043.
- Schnittler M, Eusemann P. Consequences of genotyping errors for estimation of clonality: a case study on *Populus euphratica* Oliv. (Salicaceae). *Evol Ecol.* 2010;24:1417–32.
- Meirmans PG, Van Tienderen PH. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Mol Ecol Notes.* 2004;4:792–4.

28. Goudet J. FSTAT, a program to estimate and test gene diversities and fixation indices ver. 2.9.3; 2001. Available from: <http://www2.unil.ch/popgen/softwares/fstat.htm>
29. Kalinowski ST. Hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes*. 2005;5:187–9.
30. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945–59.
31. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*. 2003;164:1567–87.
32. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes*. 2007;7:574–8.
33. Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour*. 2009;9:1322–32.
34. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*. 2005;14:2611–20.
35. Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes*. 2004;4:137–8.
36. Beaumont MA, Zhang W, Balding DJ. Approximate Bayesian computation in population genetics. *Genetics*. 2002;162:2025.
37. Cornuet J-M, Pudlo P, Veysier J, Dehne-Garcia A, Gautier M, Leblois R, Marin J-M, Estoup A. DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*. 2014;30:1187–9.
38. Li ZJ, Liu JP, Yu J, Zhou ZL. Investigation on the characteristics of biology and ecology of *Populus euphratica* and *Populus pruinosa*. *Acta Bot Boreali-Occident Sin*. 2003;23:1292–6.
39. Beerli P. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics*. 2006;22:341–5.
40. Beerli P, Felsenstein J. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proc Natl Acad Sci U S A*. 2001;98:4563–8.
41. Wilson GA, Rannala B. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*. 2003;163:1177–91.
42. Meirmans PG. Nonconvergence in Bayesian estimation of migration rates. *Mol Ecol Resour*. 2014;14:726–33.
43. Excoffier L, Laval G, Schneider S. Arlequin: an integrated software package for population genetics data analysis ver. 3.5.2.2; 2015. Available from: <http://cmpg.unibe.ch/software/arlequin35/>
44. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 1999;16:37–48.
45. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol*. 2012;29:1969–73.
46. Huang DI, Hefer CA, Kolosova N, Douglas CJ, Cronk QC. Whole plastome sequencing reveals deep plastid divergence and cytonuclear discordance between closely related balsam poplars, *Populus balsamifera* and *P. trichocarpa* (Salicaceae). *New Phytol*. 2014;204:693–703.
47. Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. *Ecol Model*. 2006;190:231–59.
48. Calagari M, Modirrahmati AR, Asadi F. Morphological variation in leaf traits of *Populus euphratica* Oliv. Natural populations. *Int J Agric Biol*. 2006;8:754–8.
49. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol*. 2005;25:1965–78.
50. Collins WD, Bitz CM, Blackmon ML, Bonan GB, Bretherton CS, Carton JA, Chang P, Doney SC, Hack JJ, Henderson TB. The community climate system model version 3 (CCSM3). *J Clim*. 2006;19:2122–43.
51. Hasumi H, Emori S. K-1 coupled gcm (miroc) description. Tokyo: Center for Climate System Research, University of Tokyo; 2004.
52. Otto-Bliesner BL, Marshall SJ, Overpeck JT, Miller GH, Hu A. Simulating Arctic climate warmth and icefield retreat in the last interglaciation. *Science*. 2006;311:1751–3.
53. Lunt DJ, Haywood AM, Schmidt GA, Salzmann U, Valdes PJ, Dowsett HJ. Earth system sensitivity inferred from Pliocene modelling and data. *Nat Geosci*. 2010;3:60–4.
54. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution*. 1984;38:1358–70.
55. Li C, Guo Z, Dupont-Nivet G. Late Cenozoic tectonic deformation across the northern foreland of the Chinese Tian Shan. *J Asian Earth Sci*. 2011;42:1066–73.
56. Zhang T, Fang X, Song C, Appel E, Wang Y. Cenozoic tectonic deformation and uplift of the south Tian Shan: implications from magnetostratigraphy and balanced cross-section restoration of the Kuqa depression. *Tectonophysics*. 2014;628:172–87.
57. Renner SS. Available data point to a 4-km-high Tibetan plateau by 40 Ma, but 100 molecular-clock papers have linked supposed recent uplift to young node ages. *J Biogeogr*. 2016;43:1479–87.
58. Yin H, Yan X, Shi Y, Qian C, Li Z, Zhang W, Wang L, Li Y, Li X, Chen G. The role of east Asian monsoon system in shaping population divergence and dynamics of a constructive desert shrub *Reaumuria soongarica*. *Sci Rep*. 2015;5:15823.
59. Zhang H-X, Zhang M-L. Genetic structure of the *Delphinium naviculare* species group tracks Pleistocene climatic oscillations in the Tianshan Mountains, arid Central Asia. *Palaeogeogr Palaeoclimatol Palaeoecol*. 2012;353:93–103.
60. Cao D, Li J, Huang Z, Baskin CC, Baskin JM, Hao P, Zhou W, Li J. Reproductive characteristics of a *Populus euphratica* population and prospects for its restoration in China. *PLoS One*. 2012;7:e39121.
61. Wiehle M, Eusemann P, Thevs N, Schnittler M. Root suckering patterns in *Populus euphratica* (Euphrates poplar, Salicaceae). *Trees-Struct Funct*. 2009;23:991–1001.
62. Wang J, Li Z, Guo Q, Ren G, Wu Y. Genetic variation within and between populations of a desert poplar (*Populus euphratica*) revealed by SSR markers. *Ann For Sci*. 2011;68:1143–9.
63. Lexer C, Fay MF, Joseph JA, Nica MS, Heinze B. Barrier to gene flow between two ecologically divergent *Populus* species, *P. alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. *Mol Ecol*. 2005;14:1045–57.
64. Rathmacher G, Niggemann M, Köhnen M, Ziegenhagen B, Bialozyt R. Short-distance gene flow in *Populus nigra* L. accounts for small-scale spatial genetic structures: implications for in situ conservation measures. *Conserv Genet*. 2010;11:1327–38.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

