

## ORIGINAL ARTICLE

# Speciation of two desert poplar species triggered by Pleistocene climatic oscillations

J Wang<sup>1</sup>, T Källman<sup>2</sup>, J Liu<sup>1</sup>, Q Guo<sup>1</sup>, Y Wu<sup>1</sup>, K Lin<sup>3,4</sup> and M Lascoux<sup>2,3</sup>

Despite the evidence that the Pleistocene climatic fluctuations have seriously affected the distribution of intraspecific diversity, less is known on its impact on interspecific divergence. In this study, we aimed to test the hypothesis that the divergence of two desert poplar species *Populus euphratica* Oliv. and *P. pruinosa* Schrenk. occurred during the Pleistocene. We sequenced 11 nuclear loci in 60 individuals from the two species to estimate the divergence time between them and to test whether gene flow occurred after species separation. Divergence time between the two species was estimated to be 0.66–1.37 million years ago (Ma), a time at which glaciation was at its maximum in China and deserts developed widely in central Asia. Isolation-with-Migration model also indicated that the two species had diverged in the presence of gene flow. We also detected evidence of selection at *GO* in *P. euphratica* and to a lesser extent at *PhyB2*. Together, these results underscore the importance of Pleistocene climate oscillations in triggering plant speciation as a result of habitats divergence.

*Heredity* (2014) **112**, 156–164; doi:10.1038/hdy.2013.87; published online 25 September 2013

**Keywords:** Pleistocene climate changes; speciation; desert poplar; selection signature

## INTRODUCTION

Accumulating evidence suggests that the Pleistocene climate oscillations, which started around 1.8 million years ago (Ma), have seriously affected the geographic distribution of species and thereby their intraspecific genetic diversity (Hewitt, 2000; Barnosky *et al.*, 2005). However, it is still unclear whether these climatic oscillations also triggered speciation events. For example, fossil evidence indicates a widespread morphological stasis throughout the Pleistocene (Morris *et al.*, 1995; Willis and Niklas, 2004). However, Pleistocene glaciations should have fragmented the continuous distributions of many species into disjunct populations, which might have promoted the production of numerous new species through allopatric speciation (Ribera and Vogler, 2003; Weir and Schluter, 2004). In addition, the new habitats created by climate changes may also have led to parapatric divergence in a few genera through adaptive evolution, especially in some climate-sensitive regions, for example, in desert regions. There are indeed a few well-examined examples of Pleistocene plant speciation (for example, Martín-Bravo *et al.*, 2010; Levens *et al.*, 2012).

Advances in population genetics inference methods based on multiple loci offer the opportunity to estimate the timing of interspecific divergence and to detect signs of adaptive evolution accompanying speciation (Hey and Nielsen, 2004). The estimated timescales based on coalescent analyses of multilocus DNA sequence data may not be as precise as estimates based on fossils, but, are good enough for the generation of a temporal hierarchy (Wakeley, 2008). In the present study, we analyzed DNA sequence variation to test the influence of Pleistocene climate oscillations on the speciation of two

desert poplars from central Asia. The species studied here are *Populus euphratica* Oliv. and *P. pruinosa* Schrenk. Both belong to sect. *Turanga* (a small poplar section with only three species) and occur commonly in central Asia as the dominant species of sparse desert forests (Eckenwalder, 1996). The natural range of *P. euphratica* extends from west China to south Morocco, whereas *P. pruinosa* is restricted to west China and adjacent regions. The third species of sect. *Turanga*, *P. ilicifolia*, occurs naturally and narrowly in Kenya and is geographically isolated and phylogenetically distant from these two species (Browicz 1966). *P. euphratica* and *P. pruinosa* are morphologically well differentiated. For example, *P. euphratica* has two types of glabrous or almost glabrous leaves. Young individuals and twigs from the basal part of old trees bear lanceolate leaves, whereas leaves from old twigs of mature trees are oval, extravagant oval, broad-triangular or dentate broad-ovate. In contrast, all leaves of *P. pruinosa* are ovate or kidney-shaped with thick hairs. In addition, the two species have different habitats: *P. euphratica* is distributed in dry deserts with deep underground water and *P. pruinosa* occurs in deserts with underground water close to the surface, near ancient or extant rivers. Despite these clear differences, these two species seem to have diverged recently because they share numerous ancestral polymorphisms at chloroplast DNA and Simple Sequence Repeat loci (Wang *et al.*, 2011). Here we used an Isolation-with-Migration (IM) model (Hey and Nielsen, 2004) to estimate the divergence time between *P. euphratica* and *P. pruinosa* using 11 unlinked nuclear loci. We also used Approximate Bayesian Computation (De Mita and Sjol, 2012) to test for departure from the standard coalescent model in each species.

<sup>1</sup>State Key Laboratory of Grassland Ecosystem, College of Life Science, Lanzhou University, Lanzhou, China; <sup>2</sup>Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden; <sup>3</sup>Laboratory of Evolutionary Genomics, CAS-MPG Partner Institute for Computational Biology, Chinese Academy of Sciences, Shanghai, China and <sup>4</sup>Graduate School of the Chinese Academy of Sciences, Beijing, China

Correspondence: Professor J Liu, Molecular Ecology Group, State Key Laboratory of Grassland Agro-Ecosystem, College of Life Science, Lanzhou University, Siyu Building, Lanzhou, Gansu 730000, China.

E-mail: liujq@nwpib.ac.cn

Received 21 February 2013; revised 7 July 2013; accepted 29 July 2013; published online 25 September 2013

## MATERIALS AND METHODS

### Sampled materials

Silica-dried leaves were collected from each of six *P. euphratica* and six *P. pruinosa* populations representative of their geographic range in west China (Figure 1; Supplementary Table S1). In each population, we collected five individuals in widely and evenly distributed sample sites and all individuals sampled were at least 500 m apart.

### Loci sampling

A total of 11 primer sets (*NHX1*, *NHX6*, *NDPK*, *GA20-ox1*, *GO*, *PhyB2*, *pC*, *HRGP*, *CPI*, *GOI* and *TR*) were designed based on expressed sequence tag sequences of *P. euphratica* (Brosche *et al.*, 2005). These loci are located on nine different chromosomes and those from the same chromosome are far apart on the *Populus trichocarpa* genome (Supplementary Tables S2, S3). In addition, we had already sequenced two chloroplast DNA fragments (*trnV* and *trnL*) and the nuclear internal transcribed spacer (ITS) region (Wang *et al.*, 2011).

### Polymerase chain reaction amplification and sequencing

PCR amplifications were performed in a 25  $\mu$ l volume containing 10–50 ng of diploid DNA, 50 mM Tris-HCl, 1.5 mM MgCl<sub>2</sub>, 0.5 mM dNTPs, 2  $\mu$ M of each primer and 1.0 units of Taq polymerase (TAKARA, Dalian, China), using a Gene-Amp PCR system 9700 DNA Thermal Cycler (Applied Biosystems, Foster City, CA, USA). The thermal profile consisted of the primary denaturing at 94 °C for 5 min, followed by 36 cycles, each of which involved 55 s at 94 °C, 1 min at the primer-specific annealing temperature and 55 s at 72 °C, and then a final extension phase of 7 min at 72 °C. Amplification products were then purified using a TIAN quick MidiPurification Kit (TIANGEN, Beijing, China) following the manufacturer's protocol. Sequencing reactions were performed with the PCR primers to cover the whole PCR segment using an ABI Prism BigDye Terminator Cycle Version 3.1 Sequencing Kit and an ABI3130 xl Genetic Analyzer (Applied Biosystems). The heterozygous PCR DNA fragments were cloned into pGEM T-easy vector and randomly selected clones were sequenced (TAKARA). Clone sequences (5–10 per allele) were compared with sequences obtained through direct sequencing and with other clone sequences for the same individual to help identify polymerase errors and PCR-mediated recombination. All sequences of the same locus were aligned with Mega 4.1 (Kumar *et al.*, 2008). All new sequences obtained in this article have been deposited in the GenBank Databases under accession numbers KC962665–KC962856.

### Nucleotide diversity

The average number of nucleotide differences per site between two sequences in a sample ( $\pi$ ) and Watterson's estimate of the population mutation rate ( $\theta_w$ ), the number of segregating sites (*S*), the number of haplotypes (*N<sub>h</sub>*) and the haplotype diversity (*H<sub>e</sub>*) were calculated using DnaSP 5.00.04 (Librado and Rozas, 2009).

### Linkage disequilibrium and recombination

The minimum number of historical recombination events (*R<sub>m</sub>*) was estimated using the four-gamete test of Hudson and Kaplan's (1985). We further investigated the decay of linkage disequilibrium (LD; McVean *et al.*, 2002) in terms of the *Z<sub>ns</sub>* statistic (Kelly, 1997) over physical pairwise distance following Remington *et al.*, (2001). Pairwise LD was estimated using the average of squared allele-frequency correlation (*r*<sup>2</sup>) between pairs of polymorphic sites. The nonlinear regression is simulated using the R statistical package (<http://www.r-project.org/>).

### Departure from the standard neutral model

The site frequency spectrum of each locus was summarized with Tajima's *D* (Tajima, 1989), Fu and Li's *D\** and *F\** (Fu and Li, 1993), Fay and Wu's *H* (Fay and Wu, 2000) as well as the *E* and *DH* test (Zeng *et al.*, 2006). Departure from the standard neutral model was tested by comparing the observed values of the summary statistics with their expected distributions. The latter were obtained through 10 000 coalescent simulations. We also used the multilocus Hudson-Kreitman-Aguadé test (HKA; Hudson *et al.*, 1987) implemented in Jody Hey's program HKA (<http://lifesci.rutgers.edu/~hey/lab>) to assess the fit of the data to the neutral equilibrium model. Besides, we repeated the HKA tests for *P. euphratica* and *P. pruinosa* separately using one sequence of *P. trichocarpa* as outgroup.

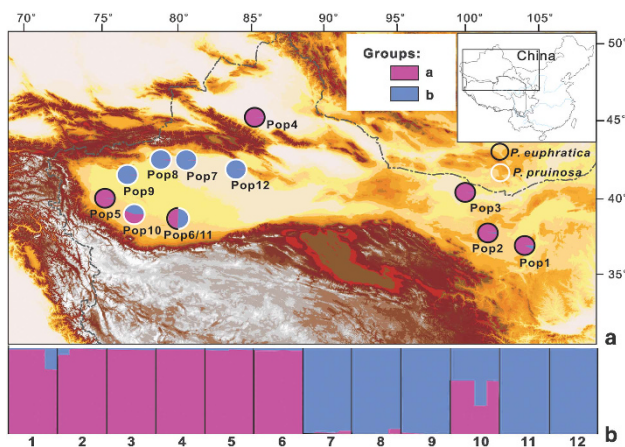
The likelihood that natural selection has occurred at individual loci was estimated using the recently developed Maximum Frequency of Derived Mutations (MFDM) test (Li, 2011), which is based on the fact that selection leads to an unbalanced gene genealogy topology. As migration may also cause unbalanced trees, we used individual sequence from the other species as a migration detector Li, 2011 to analyze this possibility. It is worth pointing out that the MFDM test is insensitive to the presence of population bottleneck or population expansion if a single Wright-Fisher population is assumed.

### Population structure

In order to quantify the extent of population genetic structure, we estimated Wright's fixation index *F<sub>ST</sub>* among populations using an analysis of molecular variance as implemented in Arlequin 3.0 program (Excoffier *et al.*, 2005). The significance of *F<sub>ST</sub>* was tested by comparing the observed value with the distribution of *F<sub>ST</sub>* after 10 000 permutations of sequences among populations. Furthermore, we used the program STRUCTURE version 2.3.2 (Hubisz *et al.*, 2009) to assess the correspondence between geographical grouping and genotypic clustering. An admixture model was used that includes linkage among polymorphisms for the ancestry model in which polymorphic sites in each locus were treated as a single linkage group and the independence of allele frequencies among populations was assumed. The number of clusters, *K* (from 1 to 12, corresponding to the number of populations), was assessed and allowance was made for the correlation of allele frequencies between clusters. Twenty runs were performed with a burn-in of 100 000 and then 500 000 iterations. The most likely number of clusters was estimated using the original method from Pritchard *et al.* (2000) and the  $\Delta K$  statistics given in Evanno *et al.* (2005).

### Within species demography

To compare the fit of the two species to different demographic models we used Approximate Bayesian Computation implemented in the egglib software package (De Mita and Siol, 2012). Three models were evaluated: (i) a standard neutral model (SNMR); (ii) a population expansion model (PEMR); and (iii) a population bottleneck model (BNMR). The first model is determined by two parameters, the population mutation rate ( $\theta$ ) and the population recombination rate ( $\rho$ ). The PEMR model has one additional parameter, the exponential growth rate ( $\alpha$ ). The BNMR is a more complex model that is



**Figure 1** Locations of sampling sites in the Northwest of China. White and black circles represent the *P. euphratica* and *P. pruinosa*, respectively. (a) Structure analysis based on 42 single nucleotide polymorphisms (SNPs). (b) The number of clusters (*K*) = 2 was showed. Bold lines distinguish populations, and blue and purple sections within each population represent the probability of assigning the individuals to each cluster. The numbers below the bars indicate the population numbers.

determined by  $\theta$ ,  $\rho$ ,  $T$  (time until end of bottleneck),  $d$  (the length of the bottleneck) and  $f$  (the size of the population during the bottleneck). The three models and associated parameters are described in more details in St Onge *et al.* (2011). We used the following wide uniform priors:  $\theta$ : 0–0.05,  $\rho$ : 0–0.5,  $\alpha$ : 0–50,  $T$ : 0–10,  $d$ : 0–10,  $f$ : 0–1 and the number of segregating sites, Tajima's  $D$  (Tajima, 1989) and the standardized version of Fay and Wu's  $H$  (Fay and Wu, 2002; Zeng *et al.*, 2006) as summary statistics. For each of the models we simulated 1 million values by randomly sampling from the prior and retained 1% of these values for linear regression of parameter values. The three models were compared using a simple rejection algorithm essentially as described in Fagundes *et al.* (2007).

### Haplotype structure

We constructed relationships between all recovered haplotypes at each nuclear locus, chloroplast DNA fragments and ITS region by the Median-Joining network. Its implementation followed the default parsimony connection limit of 95%, with indels considered as single mutation events and coded as a fifth state.

### Between-species divergence

Three approaches were used to obtain the estimate of divergence between the two species.

First, we used the log-likelihood statistics of nested models, which can be assessed using a  $\chi^2$  test to detect whether the data are consistent with a strict allopatric speciation model (Hey and Nielsen, 2007). Once runs had converged, we tested for the fit of the data to simpler demographic models by the 'Load-Trees' mode of IMA2. We created three replicate data sets that each contained a randomly chosen subset of five loci and of two populations of different distances. Then, parameters including current and ancestral effective population sizes, divergence time, long-term rates of effective migration and the ancestral population splitting were estimated using the program IMA2 based on the IM model. In cases when recombination within loci was detected, we used the IMgc program to extract maximally informative blocks of non-recombining sequences for individual markers (Woerner *et al.*, 2007).

Based on the HKY mutation model, three independent (different random number seeds) runs of at least 2 to 3 million steps following a 100 000 steps burn-in were used to ensure convergence. All runs involved 10 independent chains, and the lowest effective sample size among the six parameters was at least 200 in each case, as recommended in the IMA2 documentation. Long, well-mixing runs were repeated at least twice with different random seed. If these independent runs generated similar posterior distributions, we assumed that the chains had converged to a stationary distribution. Upper bounds of the prior distributions for each parameter were set based on the results of a preliminary run. Maxima of posterior distributions for each parameter were well under the upper bounds of the prior distributions for all three runs. Maximum-likelihood estimates and 95% highest posterior density ranges for each parameter were converted to meaningful biological quantities using a mutation rate estimate of  $2.5 \times 10^{-9}$  synonymous substitutions/site/year (Tuskan *et al.*, 2006; Invarsson 2008) and the generation time of 10 (8–15) years (Li *et al.*, 2003). Because our data include nonsynonymous and noncoding sites as well as synonymous ones, the ratio of the average total genetic divergence to the average synonymous genetic divergence ( $K_{\text{total}}/K_S$ ) was used to obtain the mutation rate ( $\mu = K_{\text{total}}/K_S \times 2.5 \times 10^{-9}$ ) (Ikeda *et al.*, 2009). The mean per-locus mutation rate was estimated to be  $2.6 \times 10^{-6}$  substitutions/locus/year, which was then used to estimate divergence time between the two species in years.

Second, we also estimated the same parameters using MIMAR (MCMC estimation of the Isolation-Migration model Allowing for Recombination; Becquet and Przeworski, 2009). Compared with IM, MIMAR takes recombination into account and is based on Hey and Wakeley (1997) summary statistics of the joint frequency spectrum. For each locus,  $S_1$  and  $S_2$  are the number of polymorphic sites unique to *P. euphratica* and *P. pruinosa*, respectively,  $S_s$  is the number of sites with shared alleles between the two species, and  $S_f$  is the number of fixed alleles in either species. The prior range for  $\theta_1$  is Uniform (0.001, 0.01), the prior range for  $T$  is uniform (0, 100 000) generations, the expected number of migrants,  $M$ , will range between (0.135 and 7.389). To test the validity of the MIMAR runs, a goodness-of-fit test was

performed using the package MIMARGOF, with simulations being carried out using the posterior distributions of the parameters estimated in MIMAR.

Third, divergence time was estimated using ITS variation between the two species, which was calculated as the average DNA sequence distance divided by twice the sequence mutation rate ( $\mu$ ). The woody perennial substitution rates ranged from  $0.38 \times 10^{-9}$  to  $7.83 \times 10^{-9}$  substitutions/site/year (mean =  $2.15 \times 10^{-9}$  subs/site/year; Kay *et al.*, 2006).

## RESULTS

### Nucleotide diversity and neutrality tests

Two haplotypes per individual were obtained for each locus, representing 60 alleles per species. The aligned sizes of the eleven loci ranged from 621 to 1863 bp, with an average size of 868 bp (9546 bp in total, Supplementary Table S2). The variation was on the average higher in *P. pruinosa* ( $\pi_t = 0.00407$ ;  $\theta_{wt} = 0.00380$ ) than in *P. euphratica* ( $\pi_t = 0.00327$ ;  $\theta_{wt} = 0.00297$ ; Table 1). Our average estimates of silent site sequence diversity are 0.00398 and 0.00506 for *P. euphratica* and *P. pruinosa*, with the highest estimates within species being 0.01442 and 0.01529, respectively.

Most tests of the standard neutral model at individual loci were nonsignificant, as summarized by the average values of the test statistics, except for *PhyB2* in *P. euphratica* that showed significant values for all tests (Table 2). We further conducted a multilocus HKA test. We did not detect any significant departure from the standard neutral model with the HKA test when considering *P. euphratica* and *P. pruinosa* ( $\chi^2 = 15.6774$ ;  $P = 0.736$ ). We compared the test statistics with a distribution generated from 10 000 coalescent simulations and did not detect significant values for both pairs of contrasts (*P. euphratica*/*P. trichocarpa*,  $\chi^2 = 2.297$ ,  $P > 0.05$ ; *P. pruinosa*/*P. trichocarpa*,  $\chi^2 = 0.663$ ,  $P > 0.05$ ).

The MFDM test detected evidence of selection at one locus: *GO*, in *P. euphratica* ( $P = 0.0339$ ; at a significance level of  $P < 0.05$ ). In addition, a low  $P$ -value was also obtained for *PhyB2* in *P. euphratica* ( $P = 0.0678$ ;  $0.05 < P < 0.07$ ). For each locus in *P. euphratica*, we arbitrarily picked one individual from *P. pruinosa* as the migration detector and one individual from *P. trichocarpa* as outgroup to build a UPGMA (Unweighted Pair Group Method with Arithmetic Mean) tree (Supplementary Figure S1). The analyses indicated that migration is not responsible for these unbalanced trees.

### Recombination and interspecific LD

Four-gamete tests indicated that the minimum number of recombination events ( $R_m$ ) in all loci for each species ranged from 0 to 3 in *P. euphratica* and from 0 to 5 in *P. pruinosa*, with the estimates being generally higher in *P. pruinosa* than in *P. euphratica* at all nuclear loci (Table 1). Supplementary Figure S2 illustrated the decline of LD with physical distance, using pooled data of all 11 loci. LD decays more rapidly to lower levels ( $r^2 < 0.2$ ) in *P. pruinosa* than in *P. euphratica*.

### Within species demographic history

In both species, the average Tajima's  $D$  values were positive (*P. euphratica* = 0.72 and *P. pruinosa* = 0.29) indicating an excess of intermediate frequency variants compared with neutral expectations, even though single loci only rarely deviated significantly. In order to identify a demographic scenario compatible with the collected sequence data three simple demographic models, a standard neutral model (SNMR), a population expansion model (PEMR) and a bottleneck scenario (BNMR), were compared. The precision in parameter estimates was in general not very high and under the BNMR,  $\theta$  was the only parameter showing a narrow, unimodal distribution (Supplementary Figure S3). The model comparisons lend

**Table 1** Nucleotide variation and haplotype diversity at 11 loci in 12 populations of *P. euphratica* and *P. pruinosa*

Species	Locus	Total					Nonsynonymous sites		Silent sites		Haplotype diversity		Recombination $R_m$
		N	L	S	$\theta_{wt}$	$\pi_t$	$\theta_{wa}$	$\pi_a$	$\theta_{sil}$	$\pi_{sil}$	$N_h$	$H_e$	
<i>P. euphratica</i>	<i>NHX1</i>	30	1863	30 (0)	0.00363	0.00473	0.00153	0.00140	0.00404	0.00537	9	0.869	1
	<i>NHX6</i>	30	885	2 (0)	0.00051	0.00111	0.00000	0.00000	0.00078	0.00168	4	0.590	0
	<i>NDPK</i>	30	636	23 (0)	0.00822	0.00983	0.00344	0.00361	0.01176	0.01442	8	0.783	2
	<i>GA20-oxl</i>	30	809	3 (0)	0.00108	0.00135	0.00700	0.00089	0.00239	0.00292	5	0.565	1
	<i>GO</i>	30	718	11 (1)	0.00335	0.00434	0.00285	0.00426	0.00352	0.00441	15	0.884	1
	<i>PhyB2</i>	30	948	21 (13)	0.00482	0.00196	0.00000	0.00000	0.00620	0.00253	13	0.759	3
	<i>pC</i>	30	701	15 (0)	0.00403	0.00377	0.00756	0.01533	0.00376	0.00288	25	0.918	2
	<i>HRGP</i>	30	940	7 (0)	0.00163	0.00211	0.00000	0.00000	0.00183	0.00237	15	0.877	3
	<i>CPI</i>	30	627	3 (0)	0.00104	0.00164	0.00134	0.00211	0.00000	0.00000	5	0.685	1
	<i>GOI</i>	30	621	4 (0)	0.00140	0.00204	0.00222	0.00288	0.00103	0.00166	6	0.664	1
	<i>TR</i>	30	798	11(0)	0.00300	0.00306	0.00000	0.00000	0.00504	0.00514	6	0.610	1
	Average	30	826	11.8	0.00297	0.00327	0.00236	0.00277	0.00370	0.00398	10	0.746	1.45
<i>P. pruinosa</i>	<i>NHX1</i>	30	1863	42 (3)	0.00510	0.00501	0.00458	0.00662	0.00521	0.00472	12	0.833	4
	<i>NHX6</i>	30	885	3 (0)	0.00077	0.00168	0.00000	0.00000	0.00118	0.00257	5	0.722	0
	<i>NDPK</i>	30	636	10 (1)	0.00340	0.00242	0.00342	0.00230	0.00342	0.00253	10	0.782	0
	<i>GA20-oxl</i>	30	809	1 (0)	0.00027	0.00039	0.00000	0.00000	0.00119	0.00172	2	0.313	0
	<i>GO</i>	30	718	29 (1)	0.00876	0.01238	0.00142	0.00220	0.01086	0.01529	19	0.756	5
	<i>PhyB2</i>	30	948	18 (2)	0.00410	0.00343	0.00103	0.00016	0.00497	0.00436	12	0.600	5
	<i>pC</i>	30	701	13 (3)	0.00401	0.00414	0.00860	0.01534	0.00365	0.00327	19	0.914	5
	<i>HRGP</i>	30	940	10(0)	0.00230	0.00223	0.00000	0.00000	0.00258	0.00250	9	0.359	2
	<i>CPI</i>	30	627	14 (4)	0.00482	0.00392	0.00489	0.00379	0.00467	0.00443	11	0.624	1
	<i>GOI</i>	30	621	11 (1)	0.00395	0.00460	0.00000	0.00000	0.00592	0.00689	6	0.591	2
	<i>TR</i>	30	798	16 (1)	0.00433	0.00454	0.00136	0.00105	0.00637	0.00692	11	0.823	5
	Average	30	826	15.2	0.00380	0.00407	0.00229	0.00285	0.00459	0.00506	11	0.665	2.64

*N*, sample size; *L*, length in base pairs; *S*, number of segregating sites (number of singletons);  $\pi$ , nucleotide diversity (Nei 1987; Nei and Li, 1979);  $\theta$ , Watterson's parameter (Watterson, 1975);  $N_h$ , number of haplotypes;  $H_e$ , Nei's haplotype diversity;  $R_m$ , minimum number of recombinant events.

support for either the SNMR or the BNMR and clearly show (as expected with positive Tajima's *D* values) that the data are not compatible with the simple growth model (Table 3).

### Population structure and differentiation

A hierarchical analysis of molecular variance approach was used to additionally quantify differentiation between the species. We observed 36.8% of the total variation among species and only 18.7% among populations within species, whereas 44.5% of the total variation was found within populations (Table 4). The average  $F_{ST}$  estimate is 0.212 for *P. pruinosa*, which was lower than the value for *P. euphratica* ( $F_{ST} = 0.367$ ). There were substantial variation across the eleven loci in levels of genetic differentiation, with the lowest estimate of  $F_{ST} = 0.322$  for *GO* and the highest estimate  $F_{ST} = 0.868$  for *NDPK* (Table 4). *Structure* indicated that the most likely number of clusters for the entire data set was  $K = 2$  (Figure 1; Supplementary Figure S4) with a clear delineation of the two species, except for one population (Pop10) of *P. pruinosa* showing a strong signal of admixture with *P. euphratica*.

### Divergences between the two species

Genealogies of the recovered haplotypes observed at each locus were constructed (Figure 2). Most of the sampled loci did not show complete lineage sorting except for *GA20-oxl* and *NDPK*. Similarly, networks for the chloroplast DNA showed no obvious divergence between the two species. In contrast, the ITS network was composed of two clusters with species-specific lineages.

The strict allopatric speciation model was rejected for the complete (all loci, all individuals), and random five loci by the 'Load-Trees' mode ( $P < 0.001$  for all tests; Supplementary Table S4). We repeated

runs of simulations with the Ima2 program and revealed unambiguous marginal posterior probability distributions of the demographic parameters for the two species. Because the loci, *GO* and *PhyB2*, are putatively under selection, we removed them from the demographic analysis. The effective sample size values for the time parameter are above 200 in the different analyses. Results are shown in Table 5 and Figure 3. Estimates of effective population sizes for each species are 97 907 for *P. euphratica* and 93 365 for *P. pruinosa*. These sizes are about two times smaller than the effective size of the ancestral population that gave rise to these species, indicating that both *P. euphratica* and *P. pruinosa* effective population sizes have decreased since the split of the two species. The marginal posterior probability distribution of the divergence parameter,  $t$ , showed a sharp peak at 2.149, which was converted into a divergence time of 0.83 Ma (0.66–1.23 Ma) with 95% highest posterior density ranged from 0.24 to 5.17 Ma (Table 5). MIMAR estimate of the divergence of the two species was of the same order of magnitude and around 1.09 Ma (0.93–1.28 Ma). Estimate of divergence time based on ITS genetic distance between the two species, and the woody perennial substitution rates ( $2.15 \times 10^{-9}$  substitutions/site/year; Kay et al., 2006), was also of the same order of magnitude with a value of 1.37 Ma. Estimates of the long-term effective migration rates between *P. euphratica* and *P. pruinosa* are asymmetric. Maximum-likelihood estimates of 2Nm from *P. pruinosa* into *P. euphratica* and from *P. euphratica* into *P. pruinosa* are 0.009 and 0.293, respectively.

### DISCUSSION

We conducted population genetic analyses of two desert poplar species, *P. euphratica* and *P. pruinosa*, based on sequence data from

**Table 2** Neutrality tests at 11 loci in 12 populations of *P. euphratica* and *P. pruinosa*. Significance was estimated with coalescent simulations under the standard neutral model

Species	Locus	D	D*	F*	H	DH (P-value)	E
<i>P. euphratica</i>	<i>NHX1</i>	1.11694	1.82313**	1.82720*	0.30436	0.8100	0.67880
	<i>NHX6</i>	2.10445*	0.75150	1.33425	0.58944	0.9524*	0.96739
	<i>NDPK</i>	0.59556	1.75322**	1.60774	-4.10232**	0.6212	4.12263**
	<i>GA20-oxl</i>	0.48047	0.99343	1.00196	0.42523	0.5880	0.00266
	<i>GO</i>	0.57136	0.84845	1.00238	0.82715	0.8179	-0.24994
	<i>PhyB2</i>	-1.86948**	-3.07873*	-3.15172*	-3.9200**	0.0024*	1.90752*
	<i>pC</i>	1.31779	1.08675	1.37129	-0.27653	0.8248	1.16160
	<i>HRGP</i>	0.75594	1.23431	1.26962	0.17823	0.6617	0.42215
	<i>CPI</i>	1.15409	0.88016	1.12335	-0.68081	0.7784	1.33782
	<i>GOI</i>	1.66880	0.99343	1.16397	0.53852	0.9078	0.74900
	<i>TR</i>	0.05518	1.43840	1.15175	-0.57880	0.3445	0.53247
	Aligned <sup>a</sup>	0.52123	1.04800	1.00814	-1.56568	0.6008	1.96714
<i>P. pruinosa</i>	<i>NHX1</i>	-0.14505	1.32503	0.98767	-1.91924*	0.3003	1.66094*
	<i>NHX6</i>	2.45780*	0.89595	1.58936	0.24396	0.9819*	1.51544
	<i>NDPK</i>	-0.79715	0.76003	0.28402	0.81788	0.8094	-1.30460
	<i>GA20-oxl</i>	0.57948	0.53080	0.63122	0.23956	0.8366	-0.09163
	<i>GO</i>	1.21346	1.55651*	1.76113*	-0.40118	0.8303	1.39400
	<i>PhyB2</i>	-0.50456	0.81914	0.42234	-1.30371	0.1806	0.72237
	<i>pC</i>	0.26980	-0.07540	-0.01920	0.57565	0.4770	-0.27448
	<i>HRGP</i>	0.02959	1.39334	1.05797	-6.56215**	0.3527	5.55248**
	<i>CPI</i>	-0.55406	-0.46366	-0.58599	-1.47884	0.1436	0.82254
	<i>GOI</i>	0.57580	0.84321	0.84457	-1.91452*	0.5942	2.06521*
	<i>TR</i>	0.05517	1.14910	0.94929	-0.57879	0.4082	0.53247
	Aligned <sup>a</sup>	0.17121	1.25730	0.99836	-1.54672	0.4448	1.63530

D, Tajima's D statistic (Tajima, 1989); D\* and F\*, Fu and Li's D\* and Fu and Li's F\* (Fu and Li 1993); and E, E test (Zeng *et al.* 2006), DH, a compound test of Tajima's D statistic and Fay and Wu's H (Zeng *et al.* 2006). Significant level: \*, 0.01 < P < 0.05; \*\*, 0.001 < P < 0.01.

<sup>a</sup>All loci are concatenated.

**Table 3** The proportion of accepted values for each of the three tested demographic models at four different tolerance levels

	Tolerance	SNMR <sup>a</sup>	PEMR <sup>b</sup>	BNMR <sup>c</sup>
<i>P. euphratica</i>	0.1	0.52	0.010	0.47
	0.05	0.53	0.006	0.47
	0.01	0.53	0.007	0.47
	0.001	0.52	0.014	0.46
<i>P. pruinosa</i>	0.1	0.52	0.012	0.47
	0.05	0.52	0.014	0.47
	0.01	0.51	0.026	0.47
	0.001	0.49	0.049	0.46

A high value shows that the majority of accepted values at a given cutoff come from this rather than from any of the two other competing models.

<sup>a</sup>A standard neutral model.

<sup>b</sup>A population expansion model.

<sup>c</sup>A population bottleneck model.

11 nuclear loci. These loci are evenly distributed across the genome (Supplementary Table S2), and should therefore provide a reasonable estimation of species-level divergence. Coalescent analyses indicated that these two poplar species diverged within the Pleistocene, at a stage when the largest glaciation occurred in Asia. These findings suggest that the increased aridity owing to the Pleistocene climate change might have triggered the recent speciation of these two poplar species in response to continuous differentiation of desert habitats. We also found that these two species diverged in the presence of gene flow and that two loci, *GO* and *PhyB2*, showed evidence of selection in *P. euphratica*.

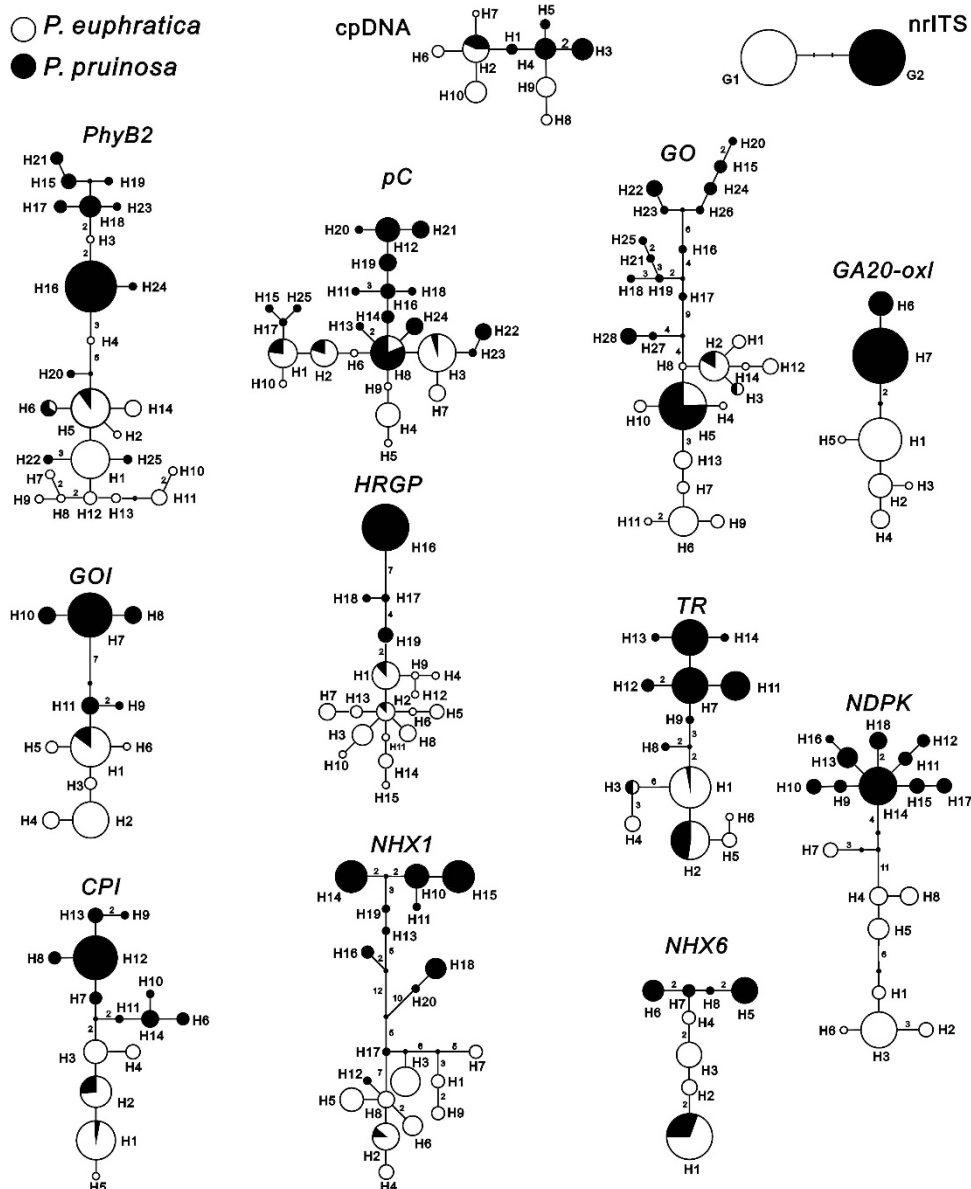
**Table 4** Summary of population differentiation ( $F_{ST}$ ) and hierarchical analysis of molecular variance (AMOVA) at 11 loci in *P. euphratica* and *P. pruinosa*

Locus	$F_{ST}$		Percentages of variation (%)			
	(Among populations within species)					
	P. euphratica	P. pruinosa	Among species	Within populations	Among populations	Among species
<i>NHX1</i>	0.219***	0.126 ***	0.692***	30.83***	8.92***	60.25***
<i>NHX6</i>	0.374***	0.134	0.520***	48.02***	12.53***	39.45***
<i>NDPK</i>	0.424***	0.310***	0.868***	13.22***	10.28***	76.50***
<i>GA20-oxl</i>	0.510***	0.076	0.703***	29.72***	16.57***	53.71***
<i>GO</i>	0.321***	0.140***	0.322***	67.79***	20.69***	11.52***
<i>PhyB2</i>	0.411***	0.149***	0.499***	50.10***	21.05***	28.85***
<i>pC</i>	0.345***	0.280***	0.424***	57.63***	25.08***	17.29***
<i>HRGP</i>	0.161***	0.268***	0.486***	51.32***	12.46***	36.23***
<i>CPI</i>	0.435***	0.290***	0.547***	45.33***	26.11***	28.56***
<i>GOI</i>	0.505***	0.200***	0.564***	43.57***	24.77***	31.67***
<i>TR</i>	0.333***	0.359***	0.480***	52.01***	27.27***	20.72***
Average	0.367	0.212	0.553	44.50	18.70	36.80

Significant level:\*\*\*: P < 0.001.

#### Gene flow and selection signature

Our IM analysis rejected a strict allopatric speciation hypothesis (Supplementary Table S4) but favored the hypothesis that the two species may have diverged in the presence of asymmetrical gene flow

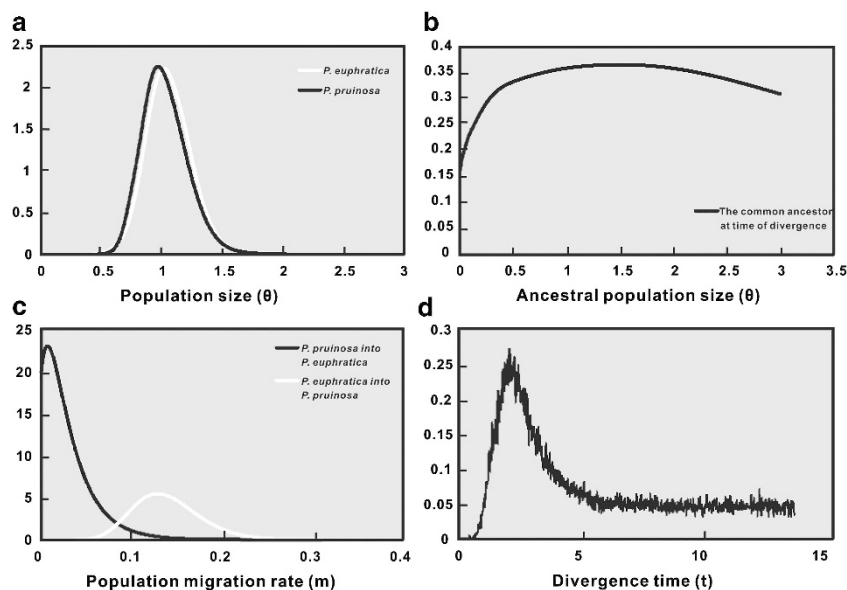


**Figure 2** Gene genealogies of the eleven nuclear loci, nrITS and two region aligned chloroplast DNA (cpDNA). Colors in the pie chart indicate the haplotype origin. The size of the pie is proportional to the haplotype frequency found in the two poplars. Branch lengths longer than one mutation step are marked on each branch.

**Table 5** MLE and the 95% HPD intervals of demographic parameters from IMA2 analyses

	$\theta_1$	$\theta_2$	$\theta_A$	$m_1$	$m_2$	$t$	$N_1$	$N_2$	$N_A$	$2N_1m_1$	$2N_2m_2$	$T_1$ (Generations)	$T_2$	$T$
MLE	1.019	0.971	1.489	0.008	0.264	2.149	97 907	93 365	143 172	0.009	0.293	66 123	123 981	82 654
HPD95Lo	0.704	0.680	0.221	0.0	0.147	1.281	67 692	58 946	19 157	0.0	0.118	19 177	35 957	23 971
HPD95Hi	1.430	1.397	2.998	0.085	0.434	13.45	137 500	134 327	288 269	0.071	0.638	413 963	776 181	517 454

Abbreviations: HPD, highest posterior density; MLE, Maximum-likelihood estimates.  
 We assume a mutation rate of  $2.5 \times 10^{-9}$  substitution/site/year and a generation time of 10 (8–15) years.  
 HPD95Lo: the lower bound of the estimated 95% HPD interval.  
 HPD95Hi: the upper bound of the estimated 95% HPD interval.  
 $\theta_1$ : population mutation rate of *P. euphratica*.  
 $\theta_2$ : population mutation rate of *P. pruinosa*.  
 $\theta_A$ : population mutation rate of ancestral population.  
 $m_1$ : population migration rate from *P. pruinosa* into *P. euphratica*.  
 $m_2$ : population migration rate from *P. euphratica* into *P. pruinosa*.  
 $t$ : time since species divergence in coalescent units.  
 $N_1$ : effective population size of *P. euphratica*.  
 $N_2$ : effective population size of *P. pruinosa*.  
 $N_A$ : effective population size of ancestral population.  
 $T$ : time since species divergence in generations.



**Figure 3** Marginal distribution of the posterior probability of six demographic parameters estimated by the IM model. (a) Effective population sizes for *P. euphratica* and *P. pruinosa*. (b) Effective population sizes for their common ancestor at the time of divergence. (c) Effective migration rates between the two species. (d) Divergence time.

(Table 5). The asymmetrical gene flow from *P. euphratica* to *P. pruinosa* is consistent with our previous study based on nuclear Simple Sequence Repeat loci (Wang *et al.*, 2011), even though it is worth pointing out that estimates of effective population sizes were vastly different between the two studies. The low effective population sizes reported in our previous study might simply reflect the inadequacy of the model assumed by *Migrate* (Beerli, 2009). In contrast to IM models that enforce a split between the two descendant species at a certain time, *Migrate* assumes an equilibrium island model. If migration is limited, the time to observe the last coalescent event may therefore be very long and smaller effective population size would therefore be required to obtain the same level of polymorphism in the two descendant species. In addition, our samples did not cover the total natural ranges of the two species and this sampling limitation may bias our estimates of effective population sizes, genetic structure and gene flow. However, the asymmetrical gene flow between the two species also agrees well with the habitat distribution of two species: *P. euphratica* occurs in the drier desert associated with deep underground water levels, whereas *P. pruinosa* occurs in damper environments within general a more limited underground water level. Therefore, *P. euphratica* can survive in the habitats of *P. pruinosa*, but the reverse seems to be more difficult, which may have accelerated the interspecific gene flow from *P. euphratica* to *P. pruinosa*.

In this study, we used various neutrality tests and the MFDM test to examine the selection signatures of each locus. The advantage of the latter is to avoid effects of demographic processes by focusing on the topology of the gene genealogy, which theoretically should not be affected by demographic events such as population expansion or bottlenecks (Li, 2011). In addition, this method considers possible gene flow and helps to minimize the number of false positives. The MFDM tests indicated that the gene, *GO*, was under selection in *P. euphratica* and also suggested the presence of selection for *PhyB2* in the same species although the signal was weaker. On the other hand, all values from Tajima's *D*, Fu and Li's *D\**, Fu and Li's *F\**, *DH* and *E* tests suggested positive selection for *PhyB2* (Table 2). The glycolate

oxidase edited by *GO* has an important role in the oxidation of glycolate to glyoxylate in plants, whereas *PhyB2* is associated with phenology (Ingvarsson *et al.*, 2006). Hence, the observed deviations at these genes do not come as a surprise, and reinforce their status as major candidate genes for further studies on adaptation in poplar species.

#### Interspecific divergence

Both IM analyses of multiple nuclear loci and ITS mutations indicate that the two species diverged during the early Pleistocene. This conclusion depends on the mutation rate and generation times adopted here. We used the same mutation rate of  $2.5 \times 10^{-9}$  per site per year for multiple nuclear loci as in previous poplar studies and the generation time of 10 (8–15) years. For ITS, we also used previous mutation calibrated for woody perennials (Kay *et al.*, 2006). The mutation rates adopted here are not direct estimates and could therefore be under- or overestimates of the true mutation rate. Similarly, estimating the average generation time remains difficult in most tree species with long lifespan (Petit and Hampe, 2006). We chose a generation time of 10 years but as the trees have a longer lifespan a higher value would also be reasonable. In a recent study of *P. euphratica*, reproductive characteristics in 20 stands from Xinjiang the mean age of the trees was 15.4 years with an average s.d. of 2.6 (Cao *et al.*, 2012). Assuming that the stands were a random sample that would imply that our estimates of generation time may not be strongly off the mark. In addition, we only sampled the co-occurring regions of two species, whereas *P. euphratica* extends further westward to Europe and southwestward to northern Africa, where *P. pruinosa* is absent. Allopatric populations may increase the overall interspecific divergence. All these uncertainties may cause some substantial error, but we believe that the estimated timescales would likely remain within the Pleistocene or the late Pliocene. These estimated stages are in line with the largest glaciation and coldest climate between 0.5 and 1.2 million years again recorded in Asian mountains. For example, the ice sheet at this stage in the Qinghai-Tibetan Plateau was five to seven

times larger than it was at later stages (Owen *et al.*, 2005). The decreased temperature caused the enlargements of the deserts in central Asia (Zheng *et al.*, 2002). This increase in desertification and diversity of desert habitats during the Pleistocene may have acted as an effective stimulus to promote the divergence of the two species.

The estimated divergence between *P. euphratica* and *P. pruinosa* is older than that between *Populus balsamifera* and *P. trichocarpa* in North America (Levsen *et al.*, 2012). However, all these estimates fall within the Pleistocene, providing further evidence for Pleistocene tree speciation in *Populus*. In contrast to the macroevolutionary stasis observed in trees since the Pliocene, these two examples, as well as those previously reported (Martín-Bravo *et al.*, 2010) highlight the importance of the Pleistocene climate oscillations for plant species diversification.

## DATA ARCHIVING

Data deposited in the Dryad repository: doi:10.5061/dryad.k70h6. Sequence data have been submitted to GenBank: accession numbers KC962665–KC962856, including the following sequence data:

CPI\_Pe: KC962665—KC962669  
CPI\_Pp: KC962670—KC962678  
GO\_Pe: KC962679—KC962693  
GO\_Pp: KC962694—KC962709  
NHX1\_Pe: KC962710—KC962717  
NHX1\_Pp: KC962718—KC962728  
NHX6\_Pe: KC962729—KC962730  
NHX6\_Pp: KC962731—KC962732  
PhyB2\_Pe: KC962733—KC962745  
PhyB2\_Pp: KC962746—KC962756  
TR\_Pe: KC962757—KC962762  
TR\_Pp: KC962763—KC962770  
GA-20oxl\_Pp: KC962771—KC962772  
GA-20oxl\_Pe: KC962773 - KC962777  
NDPK\_Pp: KC962778—KC962787  
NDPK\_Pe: KC962788—KC962798  
pC\_Pp: KC962799—KC962813  
pC\_Pe: KC962814—KC962823  
GOL\_Pe: KC962824—KC962829  
GOL\_Pp: KC962830—KC962834  
HRGP\_Pe: KC962835—KC962849  
HRGP\_Pp: KC962850—KC962856

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

Financial support was provided by the Ministry of Science and Technology (National Key Project for Basic Research 2012CB114504 and the National High Technology Research and Development Program of China No. 2013AA102605), the Fundamental Research Funds for the Central Universities (lzujbky-2009-k05) and the International Collaboration 111 Projects of China.

- Brosche M, Vinocur B, Alatalo ER, Lamminmaki A, Teichmann T, Ottow EA *et al.* (2005). Gene expression and metabolite profiling of *Populus euphratica* growing in the Negev desert. *Genome Biol* **6**: R101.
- Browicz K (1966). *Populus ilicifolia* (Engler) Rouleau and its taxonomic position. *Acta Soc Bot Pol* **35**: 325–335.
- De Mita S, Sio M (2012). EggLib: processing, analysis and simulation tools for population genetics and genomics. *BMC Genet* **13**: 27.
- Cao D, Li J, Huang Z, Baskin CC, Baskin JM, Hao P *et al.* (2012). Reproductive characteristics of a *Populus euphratica* population and prospects for its restoration in China. *PLoS One* **7**: e39121.
- Eckenwalder JE (1996). Systematics and evolution of *Populus*. In Stettler RF, Bradshaw HD, Heilman PE, Hinckley TM (eds) *Biology of Populus and its Implications for Management and Conservation*. NRC Research Press/National Research Council of Canada/Ottawa, Ontario, Canada, pp 7–32.
- Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol* **14**: 2611–2620.
- Excoffier L, Laval G, Schneider S (2005). Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol Bioinform Online* **1**: 47–50.
- Fagundes NJR, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL *et al.* (2007). Statistical evaluation of alternative models of human evolution. *Proc Natl Acad Sci USA* **104**: 17614–17619.
- Fay JC, Wu CI (2000). Hitchhiking under positive Darwinian selection. *Genetics* **131**: 479–491.
- Fay JC, Wu CI (2002). The neutral theory in the genomic era. *Curr Opin Genet Dev* **6**: 642–646.
- Fu YX, Li WH (1993). Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.
- Hewitt GM (2000). The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907–913.
- Hey J, Nielsen R (2004). Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**: 747–760.
- Hey J, Nielsen R (2007). Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc Natl Acad Sci USA* **104**: 2785–2790.
- Hey J, Wakeley J (1997). A coalescent estimator of the population recombination rate. *Genetics* **145**: 833–846.
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009). Inferring weak population structure with the assistance of sample group information. *Mol Ecol Res* **9**: 1322–1332.
- Hudson RR, Kaplan NL (1985). Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**: 147–164.
- Hudson RR, Kreitman M, Aguade M (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**: 153–159.
- Ikeda H, Fujii N, Setoguchi H (2009). Application of the isolation with migration model demonstrates the Pleistocene origin of geographic differentiation in *Cardamine nipponica* (Brassicaceae), an endemic Japanese alpine plant. *Mol Biol Evol* **26**: 2207–2216.
- Ingvansson PK, Garcia MV, Hall D, Luquez V, Jansson S (2006). Clinal variation in *PhyB2*, a candidate gene for day-length-induced growth cessation and bud set, across a latitudinal gradient in European aspen (*Populus tremula*). *Genetics* **172**: 1845–1853.
- Invarsson PK (2008). Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. *Genetics* **180**: 329–340.
- Kay KM, Whittall JB, Hodges SA (2006). A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. *BMC Evol Biol* **6**: 36.
- Kelly JK (1997). A test of neutrality based on interlocus associations. *Genetics* **146**: 1197–1206.
- Kumar S, Nei M, Dudley J, Tamura K (2008). MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequence. *Brief Bioinform* **9**: 299–306.
- Levsen ND, Tiffin P, Olson MS (2012). Pleistocene speciation in the genus *Populus* (Salicaceae). *Syst Biol* **61**: 401–412.
- Li HP (2011). A new test for detecting recent positive selection that is free from the confounding impacts of demography. *Mol Biol Evol* **28**: 365–375.
- Li ZJ, Liu JP, Yu J, Zhou ZL (2003). Investigation on the characteristics of biology and ecology of *Populus euphratica* and *Populus pruinosa* (in Chinese). *Acta bot Boreali-Occident Sin* **5**: 1–3.
- Librado P, Rozas J (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.
- Martín-Bravo S, Valcárcel V, Vargas P, Luceño M (2010). Geographical speciation related to Pleistocene range shifts in the western Mediterranean mountains (*Reseda* sect. *Glaucosceda*, Resedaceae). *Taxon* **59**: 466–482.
- McVean G, Awadalla P, Fearnhead P (2002). A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**: 1231–1241.
- Morris PJ, Ivany LC, Schopf KM, Brett CE (1995). The challenge of paleoecological stasis: reassessing sources of evolutionary stability. *Proc Natl Acad Sci USA* **92**: 11269–11273.
- Nei M (1987). *Molecular Evolutionary Genetics*. Columbia University Press: New York, NY, USA.
- Nei M, Li WH (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* **76**: 5269–5273.
- Owen LA, Finkel RC, Barnard PL, Ma HZ, Asahi K, Caffee MW *et al.* (2005). Climatic and topographic controls on the style and timing of Late Quaternary glaciation throughout

- Barnosky AD, Carrasco MA, Davis EB (2005). The impact of the species-area relationship on estimates of paleodiversity. *PLoS Biol* **3**: e266.
- Becquet C, Przeworski M (2009). Learning about modes of speciation by computation approaches. *Evolution* **63**: 2547–2562.
- Beerli P (2009). How to use migrate or why are markov chain monte carlo programs difficult to use? In: Bertorelle G, Bruford MW, Hauffe HC, Rizzoli A, Vernesi C (eds) *Population Genetics for Animal Conservation*. Cambridge University Press: Cambridge, UK. Vol 17, pp 42–79.



- Tibet and the Himalaya defined by 10Be cosmogenic radionuclide surface exposure dating. *Quaternary Science Rev* **24**: 1391–1411.
- Petit RJ, Hampe A (2006). Some evolutionary consequences of being a tree. *Ann Rev Ecol Evol Syst* **37**: 187–214.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J *et al.* (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci USA* **98**: 11479–11484.
- Ribera I, Vogler AP (2003). Speciation of Iberian diving beetles in Pleistocene refugia (*Coleoptera, Dytiscidae*). *Mol Ecol* **13**: 179–193.
- St Onge KR, Källman T, Slotte T, Lascoux M, Palmé AE (2011). Contrasting demographic history and population structure in *Capsella rubella* and *Capsella grandiflora*, two closely related species with different mating systems. *Mol Ecol* **20**: 3306–3320.
- Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U *et al.* (2006). The genome of the black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**: 1596–1604.
- Wakeley J (2008). Coalescent theory: an introduction. Greenwood Village (CO): Roberts & Company Publishers.
- Wang J, Wu Y, Ren G, Guo Q, Liu J, Lascoux M (2011). Genetic differentiation and delimitation between ecologically diverged *Populus euphratica* and *P. pruinosa*. *PLoS One* **6**: e26530.
- Watterson GA (1975). On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* **7**: 256–276.
- Weir JT, Schluter D (2004). Ice sheets promote speciation in boreal birds. *Philos Trans R Soc* **271**: 1881–1887.
- Willis KJ, Niklas KJ (2004). The role of quaternary environmental change in plant macroevolution: the exception or the rule? *Philos Trans R Soc B* **359**: 159–172.
- Woerner AE, Cox MP, Hammer MF (2007). Recombination-filtered genomic datasets by information maximization. *Bioinformatics* **23**: 1851–1853.
- Zeng K, Fu YX, Shi SH, Wu CI (2006). Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* **174**: 1431–1439.
- Zheng HB, Chen HZ, Cao JJ (2002). Palaeoenvironmental implication of the Plio-pleistocene loess deposits in southern Tarim Basin. *Chinese Sci Bull* **47**: 700–704.

Supplementary Information accompanies this paper on Heredity website (<http://www.nature.com/hdy>)