

**PHYLOGEOGRAPHY REVEALS GEOGRAPHIC AND ENVIRONMENTAL
FACTORS DRIVING GENETIC DIFFERENTIATION OF *POPULUS* SECT.
TURANGA IN NORTHWEST CHINA**

Zhongshuai Gai^{1,2,4}, Juntuan Zhai^{1,2,4}, Xiangxiang Chen^{1,2,4}, Peipei Jiao^{1,2,4},
Shanhe Zhang^{1,2,4}, Jianhao Sun^{1,2,4}, Rui Qin³, Hong Liu³, Zhihua Wu^{3*}, Zhijun Li^{1,2,4*}

¹ Key Laboratory of Biological Resource Protection and Utilization of Tarim Basin,
Xinjiang Production and Construction Group, Alar 843300, China

² College of Life Sciences, Tarim University, Alar 843300, China

³ Hubei Provincial Key Laboratory for Protection and Application of Special Plant
Germplasm in Wuling Area of China, College of Life Sciences, South-Central University for
Nationalities, Wuhan, China

⁴ Desert Poplar Research Center of Tarim University, Alar 843300, China

*CORRESPONDENCE:

Zhijun Li lizhijun0202@126.com

Zhihua Wu zhwu@scuec.edu.cn

ABSTRACT

Populus sect. Turanga (hereafter referred to as “*Populus*”), including *Populus euphratica* and *Populus pruinosa*, are the predominant tree species in desert riparian forests in northwestern China. These trees play key roles in maintaining ecosystem balance, curbing desertification, and protecting biodiversity. However, the distribution area of *Populus* forests has been severely diminished and degraded in recent years due to increased habitat destruction and human activity. Understanding the genetic diversity among *Populus* individuals and populations is essential for designing conservation strategies, but comprehensive studies of their genetic diversity in northwest China are lacking. Here, we assessed the population structures and genetic diversity of 1,620 samples from 85 natural populations of *Populus* (59 *P. euphratica* and 26 *P. pruinosa* populations) covering all of [northwestern](#) China using 120 SNP markers. Analysis of population structure revealed significant differentiation between these two-sister species and indicated that strong geographical distribution patterns, a geographical barrier, and environmental heterogeneity shaped the extant genetic patterns of *Populus*. Both *P. euphratica* and *P. pruinosa* populations in southern Xinjiang had higher genetic diversity than populations in other clades, perhaps contributing to local geographic structure and strong gene flow. Analysis of molecular variance (AMOVA) identified 15% variance among and 85% variance within subpopulations. Mantel tests suggested that the genetic variation among *P. euphratica* and *P. pruinosa* populations could be explained by both geographical and environmental distance. The genetic diversity of *P. euphratica* showed a significant negative correlation with latitude and longitude and a positive correlation with various environmental factors, such as precipitation of warmest quarter and driest month, temperature seasonality, and annual mean temperature. These findings provide insights into

how the genetic differentiation of endangered *Populus* species was driven by geographical and environmental factors, which should be helpful for designing strategies to protect these genetic resources in the future.

Keywords: *Populus euphratica*, *Populus pruinosa*, population structure, genetic diversity, geographical and environmental factors

INTRODUCTION

Populus euphratica Oliv. and *Populus pruinosa* Schrenk are sister species in the section Turanga (*Populus* sect. Turanga, hereafter referred to as “*Populus*”). Both species are predominant trees in the extremely arid desert areas in northwestern China and play important roles in maintaining ecosystem balance and protecting biodiversity (Lang et al., 2015; Zheng et al., 2016). *P. euphratica* and *P. pruinosa* diverged during the Pleistocene era due to climate oscillations and were further isolated by the barrier imposed by the Tianshan Mountains (Wang et al., 2014). Along with the hitchhiking of incidental variations, ancient genetic polymorphisms are thought to have driven the speciation of *P. euphratica* and *P. pruinosa* (Ma et al., 2018). Whereas *P. pruinosa* is mainly restricted to Xinjiang province in China and the adjacent countries and regions (Wang, 1995), the natural distribution areas of *P. euphratica* range from western China and the Middle East to North Africa and southwestern Europe (Ma et al., 2018). Approximately 61% of *P. euphratica* forests worldwide are located in China, with 91.1% of those in Xinjiang province. Alarmingly, the distribution area of *P. euphratica* forests has been seriously diminished or degraded in recent years due to increased habitat destruction and human activity. The protection of *Populus* forests in northwestern China is an extremely urgent issue. One effective strategy for protection of a plant species is to identify priority protection areas or conservation units (Zhang et al., 2020). Assessing the genetic variations in *Populus* populations would be beneficial for the conservation and utilization of the genetic resources of these species and could provide useful basic data on *Populus* germplasm to facilitate breeding.

Studies of genetic diversity and population structure are crucial for exploring natural selection, adaptive evolution, and the genetic relationships of populations within or among *P. euphratica* and *P. pruinosa*. However, although several molecular population genetics studies have been performed on *P. euphratica* and *P. pruinosa*, these studies led to opposite conclusions about their genetic diversity (Ma et al., 2018; Wang et al., 2014; Wang et al., 2011; Zeng et al., 2018). For example, a preliminary investigation using a relatively small number of nuclear and chloroplast DNA markers to investigate the genetic diversity of *P. euphratica* populations (552 *P. euphratica* individuals of 33 natural populations) indicated that this was lower in southern Xinjiang (SX) than in northern Xinjiang (NX) (Zeng et al., 2018). However, whole-genome resequencing of 252 *P. euphratica* individuals from 27 natural populations showed the reverse (Jia et al., 2020). This inconsistency might be attributed to the relatively limited numbers of markers, individuals, or populations studied in narrow geographical areas (Eusemann et al., 2013; Wang et al., 2014; Wu et al., 2008; Xu et

al., 2013), or the relatively less advanced methods utilized (e.g., AFLP, RAPD, and SRAP) (Kansu and Kaya, 2020; Saito et al., 2002; Vonlanthen, 2010; Wang et al., 2011).

The development of molecular markers at the whole-genome level is becoming increasingly accurate and efficient (Verma et al., 2015). Genome-wide single-nucleotide polymorphism (SNP) markers are widely used to assess the genotypes and genetic relationships of various populations (Lu et al., 2020; Zhang et al., 2020). Genotyping-in-Thousands by sequencing (GT-seq) has proven to be an efficient SNP genotyping technology (Campbell, 2015) and has been widely applied in many fields of research (Bootsma et al., 2020; McKinney et al., 2020; Powell and Campbell, 2020; Schmidt et al., 2020). To better understand the genetic diversity and population differentiation of *P. euphratica* and *P. pruinosa* to facilitate further conservation and breeding efforts, an analysis of their population genetics using larger-scale samples and more advanced molecular markers is needed.

The natural ranges of *P. euphratica* and *P. pruinosa* include different niches that show discontinuous geographical distribution. The barrier represented by the Tianshan Mountains, a major mountain system and biodiversity hotspot, has impeded gene flow between SX and NX in *Populus* populations (Zeng et al., 2018). The geographical differences caused by environmental heterogeneity and climatic fluctuations could drive plant-specific or intraspecific genetic divergence and distribution patterns (Muellner-Riehl, 2019). Geographical distance can also affect the gene flow from one population to another, while environmental factors can influence genetic diversity and help fix a population in a specific ecological niche (Liu et al., 2013). However, there is no report on how the genetic distance and differentiation of *P. euphratica* and *P. pruinosa* were influenced by geographical and environmental factors. Elucidating the relationship between genetic diversity and geographical or environmental heterogeneity will help explain the discontinuous distribution pattern of *Populus* in northwestern China.

In this study, we collected the leaves of 1,620 accessions from 63 natural populations throughout the natural distribution range of *P. euphratica* (1,183 individuals from 59 natural populations) and *P. pruinosa* (437 individuals from 26 natural populations) in [northwest](#) China. Based on our *de novo* assembled reference genome and published resequencing data, we developed 120 high-quality SNPs with $MAF \geq 0.4$ for further genetic analysis. Our objectives were to (1) determine the optimal population structures of *P. euphratica* and *P. pruinosa*; (2) evaluate genetic diversity and genetic differentiation between and within the subpopulations; and (3) assess the relationships between population-level genetic diversity and geographical as well as environmental factors. Our findings provide new insights into the genetic differentiation of *P. euphratica* and *P. pruinosa* and lay a genetic foundation for the construction of priority conservation areas of *Populus*.

MATERIALS AND METHODS

Sample Collection and DNA Extraction

Leaves were sampled from 1,183 *P. euphratica* individuals from 59 natural populations in northwest China, including 27 populations from southern Xinjiang (SX), 23

from northern Xinjiang (NX), and 9 total from Gansu (GS), Qinghai (QH), Ningxia (NiX), and Inner Mongolia (NMG) provinces (Supplementary Table S1). Leaves were sampled from 437 *P. pruinosa* individuals from 26 natural populations in Xinjiang province (23 from southern Xinjiang, 3 from northern Xinjiang), 22 of which co-occurred with *P. euphratica*. To reduce the impact of clonal ramets, all sampled individuals from each population were located at least 100 m apart. The leaves were dried with silica gel in the field. Total DNA was extracted from the leaves and purified using a modified CTAB method (Allen et al., 2006). The integrity of the DNA was evaluated by 0.8% agarose gel electrophoresis, and DNA quality was evaluated based on the A_{260}/A_{280} absorbance ratio of the samples.

SNP Marker Development

Marker development and sequencing were carried out using the GT-seq method (Campbell, 2015). In brief, 120 SNP markers were selected and developed for the *P. euphratica* samples based on our reference genome and resequencing data from *P. euphratica* (Jia et al., 2020). These SNPs, which are uniformly distributed on the 19 *P. euphratica* chromosomes, were retained according to the following criteria (Supplementary Figure S1): (i) SNP interval > 50 kb; (ii) minor allele frequency (MAF) ≥ 0.4 ; (iii) heterozygosity $\leq 25\%$; (iv) missing rates ≤ 0.25 . Primer 3 software (version 2.5.0) was used to design amplification primers for each target site according to the following criteria: (i) the primers ranged from 17 to 32 bases in length; (ii) melting temperature (T_m) values ranged from 60 °C to 64 °C, with an optimal value of 62 °C; (iii) product size was ≤ 500 bp; (iv) sequenced reads had to be able to cover target sites. Primers for each target site that could specifically amplify the target were obtained (Supplementary Table S2). Libraries were constructed and sequenced (PE150) on the Illumina HiSeq 2500 platform.

Sequencing Data Analysis

Reliable clean reads were obtained via quality control and filtering of raw reads. Quality control was performed based on strict criteria to remove the following types of reads: (i) adaptor sequences of reads were removed using Cutadapt software (version 1.13); (ii) low-quality bases were removed using Trimmomatic software (Version 0.36); (iii) reads ≤ 50 bp long were excluded. The reliable clean reads were aligned to the *P. euphratica* reference genome using the BWA-MEM algorithm (version 0.7.15-r1140). After alignment, GATK (version 3.7) (DePristo et al., 2011) and SAMtools (version 1.3.1) (Li and Durbin, 2009) were used for SNP calling. Genotypes at SNP loci were analyzed using an in-house Python script.

Analysis of Population Structure

Population structure analysis was implemented using STRUCTURE v2.3.4 (Earl and vonHoldt, 2011). The K -value was used to evaluate the number of clusters based on the genotyping data; the number of genetic clades was predefined using 10,000 iterations for each run. The optimal K -value was determined using Structure Harvester (Earl and VonHoldt, 2012). Principal component analysis (PCA) was carried out using GCTA (Yang et al., 2011). A neighbor-joining (NJ) tree was constructed and visualized using MEGA 7 (Tamura et al., 2013).

Analysis of Molecular Variance and Genetic Diversity

Analysis of Molecular Variance (AMOVA) was performed and Nei's genetic distance was calculated using GenAlEx 6.503 based on the number of subpopulations (Peakall and Smouse, 2012). Pairwise population differentiation (F_{ST}) was calculated. Multiple indicators of genetic diversity from each population, such as Shannon's information index (I), observed heterozygosity (H_o), expected heterozygosity (H_e), and polymorphism information content (PIC), were obtained using GenAlEx v6.503.

Analysis of Genetic Isolation

To study the relationships of isolation-by-environment versus isolation-by-distance based on the genetic differentiation of *P. euphratica* and *P. pruinosa*, 19 bioclimatic variables with a resolution of 2.5 arc-min from the WorldClim Database (<http://www.worldclim.org>) were obtained (Fick and Hijmans, 2017). Their correlation coefficients were then calculated, and the seven least correlated bioclimatic variables (Spearman's $r < 0.8$) were retained (Supplementary Tables S3 and S4). The correlation coefficients between pairwise genetic differentiation and the bioclimatic distance matrix, as well as the geographical distance matrix, were calculated. Mantel tests were performed with 1,000 permutations based on Genepop (<https://genepop.curtin.edu.au/>) (Rousset, 2008).

To examine the relationships between genetic diversity at the population level and geographical factors, as well as environmental factors according to a previous study (Zhang et al., 2020), the genetic diversity of each population was represented by expected heterozygosity (H_e). Latitude and longitude were considered to be the main geographical factors affecting the *P. euphratica* and *P. pruinosa* populations. All seven retained bioclimatic variables were environmental factors. PCA was used to analyze the climate variables of the sample sites, and the absolute values from standardized PC1 scores were obtained based on correlation analyses using R version 3.5.0.

RESULTS

Population Structure and Genetic Relationships

To comprehensively investigate the genetic diversity of *P. euphratica* (1,183 individuals from 59 natural populations) and *P. pruinosa* (437 individuals from 26 natural populations), we sampled 1,620 individuals from 85 natural *Populus* populations covering the distribution regions throughout northwestern China (Figure 1). The number and distribution area of *P. euphratica* are larger in SX than in NX, and *P. pruinosa* is primarily distributed in SX. The habitat regions of *P. euphratica* showed discontinuous distribution on both banks of inland rivers in the arid regions (e.g., the Tarim River, [Yarkant River](#), and Hotan River of southern Xinjiang and the Ili River, Manas River, and [Eerqisi River](#) of northern Xinjiang). In addition, for *P. euphratica*, compared to previous studies, we collected samples from the easternmost population (SZWQ) of China for the first time and added samples from Ningxia Province (ZW). For *P. pruinosa*, compared to previous studies, samples from three populations of *P. pruinosa* (CX, KKDL, and NLKX) were collected for the first time in northern Xinjiang (Figure 1 and Supplementary Table S1). In summary,

1,620 samples with high coverage from northwestern China were collected to investigate the genetic diversity and population structures of *P. euphratica* and *P. pruinosa*.

We then carried out population structure analysis based on our genotypic data (Figure 2). When $K = 2$, all individuals were clearly subdivided into two species-specific clusters, highlighting the significant differentiation of *P. euphratica* and *P. pruinosa* at the species level. Notably, a few *P. euphratica* individuals were mixed with the *P. pruinosa* lineage (e.g., MGTX, YPHX, BCX, ALE, YTX), implying the presence of recent gene flow in the two species. When $K = 3$, the *P. euphratica* populations were further divided into southern Xinjiang, northern Xinjiang, and mixed clusters (including GS, NMG, NiX, and QH populations). When $K = 4$, the *P. euphratica* populations were divided into four distinct clades that demonstrated strong geographical distribution patterns (NX, SX, GNM, and QH, with the GNM clade containing individuals from Gansu, Ningxia and Inner Mongolia individuals).

PCA and neighbor-joining phylogenetic analysis [confirmed](#) our findings about population structure. PC1 mainly explained the variation between *P. pruinosa* and *P. euphratica*, and intermediate individuals (SX1) were primarily from the mixed forests of *P. pruinosa* and *P. euphratica* in southern Xinjiang (e.g., MGTX, ALE, and LPX) (Figure 3 and Supplementary Table S5). The presence of these intermediate individuals provides further evidence of gene flow between the two species. PC2 strongly explained the differentiation among the four clades (SX, NX, QH, and GNM) of *P. euphratica*. The QH cluster demonstrated distinct genetic variation compared to the other clades, which might be related to the high altitude (2,709–2,936 m) of this region. The phylogenetic tree, which was constructed to represent the genetic distances between these populations, revealed distinctly different evolutionary relationships among the populations (Figure 4). All three complementary methods (STRUCTURE, PCA, and NJ tree analysis) indicated that distinct genetic differences existed with different geographical distribution patterns at the population level and that the two sister species (*P. euphratica* and *P. pruinosa*) showed significant differentiation at the species level regardless of geographical distance.

Genetic Diversity of *P. euphratica* and *P. pruinosa* Populations

Based on the population structures of *P. euphratica* and *P. pruinosa* (Figure 2), we calculated the four genetic indices (H_e , H_o , I , and PIC) for each clade and population. The genetic diversity values revealed higher diversity for *P. pruinosa* ($H_e = 0.256$) than for the four clades of *P. euphratica*. For *P. euphratica*, the SX clade exhibited the highest genetic diversity ($H_e = 0.216$), followed by the GNM clade ($H_e = 0.212$) and the NX clade ($H_e = 0.206$), while the QH clade had the lowest genetic diversity ($H_e = 0.112$) (Table 1).

In addition, we calculated the genetic diversity of each population within *P. euphratica* and *P. pruinosa* (Supplementary Tables S6 and S7). Among *P. euphratica* populations, the MGTX population (from the KS region of southern Xinjiang) had the highest genetic diversity ($H_e = 0.29$) and the TLX population (from the TC region of northern Xinjiang) had the lowest genetic diversity ($H_e = 0.11$). [On the whole](#), the populations of southern Xinjiang had relatively high genetic diversity (e.g., MGTX, ALE,

YPHX, LPX, PSX, and YTX), followed by one population from Inner Mongolia (EJN) and those from Ningxia (ZW) and Gansu (GZX and JTX), while the populations of northern Xinjiang (e.g., TLX, MLX, FYX, HBHX, and ALSK) and QH (GEM), and the second population from Inner Mongolia (SZWQ), had relatively low genetic diversity (Supplementary Table S6). Among *P. pruinosa* populations, the YTX population (HT region of southern Xinjiang) had the highest genetic diversity ($H_e = 0.28$), while the YLX population (BaZ region of southern Xinjiang) had the lowest ($H_e = 0.12$). The three *P. pruinosa* populations (CX, KKDL, and NLKX) found only in northern Xinjiang had relatively low genetic diversity (Supplementary Table S7). Overall, the populations of *Populus* in southern Xinjiang had higher genetic diversity than those of other clades. Of these, the MGTX, YTX, ALE, BCX, and KPX populations had relatively high genetic diversity, whereas the KCX, KKDL, AKTX, NLKX, and YLX populations had lower genetic diversity.

Genetic Differentiation of the Populations

The genetic divergence (F_{ST} from 0.118 to 0.160) between *P. euphratica* and *P. pruinosa* was high (Table 2). Of the four clades of *P. euphratica*, we detected high genetic differentiation between QH and GNM ($F_{ST} = 0.064$), moderate differentiation between NX and SX ($F_{ST} = 0.044$) and between NX and QH ($F_{ST} = 0.044$), and lower genetic differentiation between SX and GNM ($F_{ST} = 0.010$).

We further analyzed the genetic differentiation within *P. euphratica* and *P. pruinosa* via pairwise comparisons of different populations. For *P. euphratica*, the genetic differentiation coefficient within populations in southern Xinjiang was low except for MGTX, ALE, and CLX, indicative of strong gene flow among populations. Detailed PCA showed that most of the MGTX (9/11) and ALE (8/11) samples were in SX1 (Supplementary Table S5), suggesting that the higher genetic diversity of MGTX and ALE might have been caused by hybridization of *P. euphratica* and *P. pruinosa*. The genetic differentiation among *P. euphratica* populations in southern Xinjiang and populations of other clades (NX, GNM, and QH) was higher, especially for MLX, FYX, TLX, MQX, SZWQ, and GEM (Supplementary Table S8). However, higher levels of genetic differentiation occurred within the northern Xinjiang population between NX and other populations (GNM and QH). Similar results were obtained regarding the genetic differentiation between the GNM and QH populations. For *P. pruinosa*, the genetic differentiation within populations in southern Xinjiang was high except for MGTX and BCX, and the genetic differentiation among populations in SX and NX was high. Notably, QMX exhibited high genetic differentiation from all other populations of *P. pruinosa* (F_{ST} from 0.180 to 0.397), which is associated with the isolated geographic location of this region (Supplementary Table S9). Overall, the populations of *Populus* had relatively high genetic differentiation between SX and NX. AMOVA revealed higher genetic variation within populations (85%, a total of 36% among individuals and 49% within individuals) than among populations (15%) of *P. euphratica*, suggesting a high level of differentiation (Table 3). Nei's genetic distance analysis revealed moderate levels of population

differentiation ($F_{ST} = 0.147$) accompanied by a high rate of gene flow between *P. euphratica* populations ($N_m = 1.447$).

Potential Distributions and Genetic Isolation

To determine whether the genetic differentiation in *Populus* is significantly associated with isolation-by-distance and/or isolation-by-environment, we conducted Mantel tests between pairwise genetic differentiation ($F_{ST} / (1 - F_{ST})$) and a geographical distance matrix, as well as a Euclidean bioclimatic distance matrix with the seven bioclimatic variables. Geographical distance (Figure 5A) and environmental distance (Figure 5B) had significant effects on genetic distance among *P. euphratica* populations. Positive correlations of genetic distance with geographical distance (Figure 5C) and environmental distance (Figure 5D) were also observed for the *P. pruinosa* populations. This finding implies that the genetic differentiation of *Populus* increased as the geographical and environmental distances between populations increased, leading to low gene flow. Also, the different correlations for *P. euphratica* and *P. pruinosa* suggest that environmental distance played a more important role in the genetic differentiation of the *P. euphratica* populations ($R^2 = 0.11$) than the *P. pruinosa* populations ($R^2 = 0.05$), while geographical distance contributed equally to the genetic differentiation of both the *P. euphratica* and *P. pruinosa* populations.

Finally, to test the effects of environmental and geographical factors on the genetic diversity (H_e) of *P. euphratica* and *P. pruinosa* at the population level, we performed correlation analysis between the genetic diversity of *Populus* and latitudinal, longitudinal, and environmental factors. H_e showed significant negative correlations with longitude and especially latitude for the *P. euphratica* populations (Figure 6A and B), suggesting that the genetic diversity of this species markedly decreased with increasing latitude and longitude. The H_e of the *P. pruinosa* populations was negatively correlated with latitude but not longitude (Figure 6D and E). Thus, the negative influence of latitude on the genetic diversity of *Populus* was more important than that of longitude. The genetic diversity values of both the *P. euphratica* and *P. pruinosa* populations showed significant positive correlations with environmental factors (Figure 6C and F), suggesting that such factors play important roles in the genetic diversity of both species.

DISCUSSION

Geographical isolation and environmental heterogeneity are considered to be the crucial drivers of allopatric variance (Binks et al., 2019; Evans et al., 2014; Jiang et al., 2018). A geographic barrier not only facilitates the local adaptation of populations but also can lead to the evolution of populations with unique genetic characteristic (Barnes et al., 2002; Zhang et al., 2020). Our results reveal distinct differentiation between *P. euphratica* populations of the NX and SX clades, suggesting that the Tianshan Mountains may limit gene flow between the two clades and that the differences in local environments and the long period of isolation facilitated their allopatric divergence. It is also thought that the Tianshan Mountains impede seed dispersal and pollen flow in these two regions (Jia et al., 2020; Zeng et al., 2018). Accumulating evidence suggests that climatic fluctuations during the Pleistocene era led to extreme drought and desert expansion in southern Xinjiang (Meng

et al., 2015), resulting in habitat fragmentation of desert plant populations. This fragmentation likely affected the geographical distributions of different species, further promoting population divergence and affecting intraspecific genetic diversity (Larmuseau et al., 2009; Ye et al., 2019). Here, Mantel tests showed that the greater the geographical distance, the higher the genetic differentiation and the weaker the gene flow between populations (Figure 5). Moreover, in the current study, we collected CX, KKDL and NLKX populations of *P. pruinosa* in northern Xinjiang for the first time, and we found that the above analyses were also applicable to *P. pruinosa*.

Genetic diversity refers to the variation of genetic characteristics in a population, which allows the population to adapt to environmental changes. The genetic diversities of each clade were ranked from high to low as follows: SX > GNM > NX > QH. Our results are similar to previous findings based on whole-genome resequencing (Jia et al., 2020) but are not quite consistent with findings based on analysis using 17 nuclear microsatellite loci from *Populus trichocarpa* (Zeng et al., 2018). Perhaps the relatively small number and the *P. trichocarpa* origin of these molecular markers limit their applicability for evaluating genetic diversity in *P. euphratica* and *P. pruinosa*.

Accumulating evidence suggests that gene flow is a creative force in evolution that functions via pollen spread or seed dispersal. Intraspecific genetic admixture can produce new allelic combinations, leading to novel genotypes and phenotypes (Hendry, 2002; Olson-Manning, 2012; Rius and Darling, 2014). Interestingly, the *P. euphratica* MGTX population showed variable leaf morphology intermediate between that of *P. euphratica* and *P. pruinosa* (Supplementary Figure S2), along with the highest genetic diversity among the *P. euphratica* populations, which could be attributed to gene flow and gene introgression between species (Supplementary Table S5). In addition, *P. euphratica* and *P. pruinosa* are dioecious, with males flowering earlier than females in each species, although the flowering periods basically overlap. However, *P. euphratica* flowers slightly earlier than *P. pruinosa*, suggesting that asymmetrical gene flow from male *P. pruinosa* to female *P. euphratica* might increase the probability of genetic variation (Wang et al., 2011; Wang, 1995). This observation likely explains why the co-occurring population of *P. euphratica* and *P. pruinosa* had higher genetic diversity than the other populations of each species.

Current assumptions suggest that palaeogeographical changes (Wang et al., 2013) or heterogeneous environments (Forester et al., 2016) might also affect the spatial genetic patterns of plants. We found here that both geographical distance and environmental heterogeneity significantly influenced the genetic variation of *P. euphratica* and *P. pruinosa*, suggesting that the genetic distance within the isolates increased with geographical distance. Changes in longitude and latitude represent the effects of geographical distance on genetic variation, which further strengthens our hypothesis that the presence of a geographical barrier facilitated the genetic differentiation of these populations. Local environmental differences and a long period of isolation facilitate allopatric divergence. For example, the environmental factors of NMG, NiX, and GS were intermediate between those of SX and NX (Jia et al., 2020), which might have contributed

to the formation of the GNM clade. Meanwhile, analysis of the effects of geographical and environmental factors on the genetic diversity of *P. euphratica* and *P. pruinosa* showed that genetic diversity decreased with increasing latitude and longitude. In particular, the genetic diversity of *Populus* is more closely related to latitude, which might be related to differences in photoperiod or temperature in different latitudes. This notion requires further study.

A recent study of environmental factors influencing the geographical distribution of *P. euphratica* showed that this is affected by precipitation of the driest month and the warmest quarter, the soil moisture content (10–40 cm underground), soil moisture around the root system, and the evapotranspiration of soil water (Guo et al., 2020). Precipitation of the driest month and the warmest quarter is an important source of soil groundwater recharge during the growing season of *P. euphratica*. Soil water evaporation limits the effective recharge of soil groundwater, and soil water content and root soil moisture are important indices of effective soil groundwater recharge. Notably, precipitation of the driest month and the warmest month was opportunely encountered flooding season in southern Xinjiang (Gai et al., 2020; Zheng et al., 2016). In general, the peak of yearly seed release (at the end of July) coincides with the annual flooding period in that region (Eusemann et al., 2013). These favorable conditions may be conducive to the reproduction of *P. euphratica* and *P. pruinosa*, perhaps explaining why southern Xinjiang owns the largest *P. euphratica* and *P. pruinosa* forests in northwest China.

The high genetic diversity of the SX populations may result from its distinct local geographic structure and strong gene flow (Shen et al., 2014). Southern Xinjiang represents the widest distribution area of *P. euphratica* globally. The high genetic diversity of *P. euphratica* populations in this region reflects the large effective population size and large amount of endemic gene resources. To better protect these genetic *Populus* resources, more effort should focus on the intensity of flood diversion and irrigation and the protection of *P. euphratica* forests in southern Xinjiang, regardless of their ecological significance or importance to genetic conservation. More attention should also be paid to the replacement and renewal of *P. euphratica* forests in northern Xinjiang. Finally, the dramatic decreases in *P. pruinosa* populations underline the need to preserve this species, which is important not only in itself but as a precious genetic resource to enrich the gene pool of *P. euphratica*. Our genetic data here indicate that a comprehensive strategy is needed to maximize the protection of desert poplar.

CONCLUSION

In this study, we explored population structure and genetic diversity among 1,620 samples covering the full distribution ranges of *P. euphratica* and *P. pruinosa* populations in [northwest](#) China. Our study found that the natural populations of *P. euphratica* in northwest China show strong geographical distribution patterns and that *P. euphratica* and *P. pruinosa* populations in southern Xinjiang have higher genetic diversity than the populations of other clades. Mantel tests suggested that both geographical and environmental distance significantly influence genetic variation in *P. euphratica* and *P.*

pruinosa: the genetic diversity of *P. euphratica* **markedly** decreases with increasing latitude and longitude and is positively correlated with environmental factors. Finally, the higher genetic diversity in southern Xinjiang may contribute to the local geographic structure and strong gene flow in these populations.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The reference genome of *Populus euphratica* is available at our *Populus* genome database (<http://118.24.202.236:20003/>). The names of the SNP repository and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/>, PRJNA738601.

AUTHOR CONTRIBUTION

ZL and ZW conceived and designed the study and provided suggestions and comments about the manuscript. ZG collected and analyzed the data and wrote the manuscript. JZ, XC, PJ, SZ, and JS collected the samples. RQ and HL provided constructive comments. ZW revised the manuscript. All authors read and approved the manuscript.

ACKNOWLEDGMENTS

This work was financially supported by the National Natural Sciences Foundation of China (31860198 and 30660018), and Innovative Team Building Plan for Key Areas of Xinjiang Production and Construction Corps (2018CB003), Xinjiang Production & Construction Corps Key Laboratory of Protection and Utilization of Biological Resources in Tarim Basin (BRZD2003). We are grateful to Dr. Hong-Xiang Zhang from the Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, for his assistance with data analysis. We are grateful to our many colleagues in forestry research from Xinjiang, Gansu, Ningxia, Qinghai, and Inner Mongolia provinces for their help during sample collection. We are grateful for technical support from Wuhan Genoseq Science and Technology Co., Ltd.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES:

1. Allen, G.C., Flores-Vergara, M.A., Krasynanski, S., Kumar, S., and Thompson, W.F. (2006). A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat Protoc* 1, 2320-2325.
2. Barnes, I., Matheus, P., Shapiro, B., Jensen, D., and Cooper, A. (2002). Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science* 295, 2267-2270.
3. Binks, R.M., Gibson, N., Ottewell, K.M., Macdonald, B., and Byrne, M. (2019). Predicting contemporary range-wide genomic variation using climatic, phylogeographic and morphological knowledge in an ancient, unglaciated landscape. *Journal of Biogeography* 46, 503-514.
4. Bootsma, M.L., Gruenthal, K.M., McKinney, G.J., Simmons, L., Miller, L., Sass, G.G., and Larson, W.A. (2020). A GT-seq panel for walleye (*Sander vitreus*) provides important insights for efficient development and implementation of amplicon panels in non-model organisms. *Mol Ecol Resour* 20, 1706-1722.
5. Campbell, N.R.H.S.A.a.N.S.R. (2015). Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Mol Ecol Resour* 15, 855-867.
6. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., *et al.* (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43, 491-498.
7. Earl, D.A., and vonHoldt, B.M. (2011). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4, 359-361.
8. Eusemann, P., Petzold, A., Thevs, N., and Schnittler, M. (2013). Growth patterns and genetic structure of *Populus euphratica* Oliv. (Salicaceae) forests in NW China – Implications for conservation and management. *Forest Ecology and Management* 297, 27-36.
9. Evans, L.M., Slavov, G.T., Rodgers-Melnick, E., Martin, J., Ranjan, P., Muchero, W., Brunner, A.M., Schackwitz, W., Gunter, L., Chen, J.G., *et al.* (2014). Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nat Genet* 46, 1089-1096.
10. Fick, S.E., and Hijmans, R.J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37, 4302-4315.
11. Forester, B.R., Jones, M.R., Joost, S., Landguth, E.L., and Lasky, J.R. (2016). Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Mol Ecol* 25, 104-120.

12. Gai, Z., Li, X., Zhai, J., Chen, X., and Li, Z. (2020). Effects of River Bank Soil Physical and Chemical Factors on *Populus Pruinosa* Schrenk Clonal Growth. *Applied Ecology and Environmental Research* 18, 4791-4806.
13. Guo, F., Xu, G., Lu, M., Meng, Y., Yuan, C., and Guo, K. (2020). Prediction of Potential Suitable Distribution Areas for *Populus euphratica* Using the MaxEnt Model. *Scientia Silvae Sinicae* 56, 184-192.
14. Hendry, A.P., Taylor ERIC B, and J DONALD Mcphall (2002). Adaptive divergence and the balance between selection and gene flow_ lake and stream stickleback in the Misty system. *Evolution* 56, 1199-1216.
15. Jia, H., Liu, G., Li, J., Zhang, J., Sun, P., Zhao, S., Zhou, X., Lu, M., and Hu, J. (2020). Genome resequencing reveals demographic history and genetic architecture of seed salinity tolerance in *Populus euphratica*. *J Exp Bot* 71, 4308-4320.
16. Jiang, X.L., An, M., Zheng, S.S., Deng, M., and Su, Z.H. (2018). Geographical isolation and environmental heterogeneity contribute to the spatial genetic patterns of *Quercus kerrii* (Fagaceae). *Heredity* 120, 219-233.
17. Kansu, Ç., and Kaya, Z. (2020). Genetic diversity of marginal populations of *Populus euphratica* Oliv. from highly fragmented river ecosystems. *Silvae Genetica* 69, 139-151.
18. Lang, P., Jeschke, M., Wommelsdorf, T., Backes, T., Lv, C., Zhang, X., and Thomas, F.M. (2015). Wood harvest by pollarding exerts long-term effects on *Populus euphratica* stands in riparian forests at the Tarim River, NW China. *Forest Ecology and Management* 353, 87-96.
19. Larmuseau, M.H.D., Van Houdt, J.K.J., Guelinckx, J., Hellemans, B., and Volckaert, F.A.M. (2009). Distributional and demographic consequences of Pleistocene climate fluctuations for a marine demersal fish in the north-eastern Atlantic. *Journal of Biogeography* 36, 1138-1151.
20. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-1760.
21. Liu, Y., Webber, S., Bowgen, K., Schmaltz, L., Bradley, K., Halvarsson, P., Abdelgadir, M., and Griesser, M. (2013). Environmental factors influence both abundance and genetic diversity in a widespread bird species. *Ecol Evol* 3, 4683-4695.
22. Lu, R.S., Chen, Y., Tamaki, I., Sakaguchi, S., Ding, Y.Q., Takahashi, D., Li, P., Isaji, Y., Chen, J., and Qiu, Y.X. (2020). Pre-quaternary diversification and glacial demographic expansions of *Cardiocrinum* (Liliaceae) in temperate forest biomes of Sino-Japanese Floristic Region. *Mol Phylogenet Evol* 143, 106693.
23. Ma, T., Wang, K., Hu, Q., Xi, Z., Wan, D., Wang, Q., Feng, J., Jiang, D., Ahani, H., Abbott, R.J., *et al.* (2018). Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. *Proc Natl Acad Sci U S A* 115, E236-E243.

24. McKinney, G.J., Seeb, J.E., Pascal, C.E., Schindler, D.E., Gilk-Baumer, S.E., and Seeb, L.W. (2020). Y-chromosome haplotypes are associated with variation in size and age at maturity in male Chinook salmon. *Evol Appl* 13, 2791-2806.
25. Meng, H.-H., Gao, X.-Y., Huang, J.-F., and Zhang, M.-L. (2015). Plant phylogeography in arid Northwest China: Retrospectives and perspectives. *Journal of Systematics and Evolution* 53, 33-46.
26. Muellner-Riehl, A.N. (2019). Mountains as Evolutionary Arenas: Patterns, Emerging Approaches, Paradigm Shifts, and Their Implications for Plant Phylogeographic Research in the Tibeto-Himalayan Region. *Front Plant Sci* 10, 195.
27. Olson-Manning, C.F.W., M. R. and Mitchell-Olds, T. (2012). Adaptive evolution: evaluating empirical support for theoretical predictions. *Nat Rev Genet* 13, 867-877.
28. Peakall, R., and Smouse, P.E. (2012). GenAEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics* 28, 2537-2539.
29. Powell, J.H., and Campbell, M.R. (2020). Contemporary genetic structure affects genetic stock identification of steelhead trout in the Snake River basin. *Ecol Evol* 10, 10520-10531.
30. Rius, M., and Darling, J.A. (2014). How important is intraspecific genetic admixture to the success of colonising populations? *Trends Ecol Evol* 29, 233-242.
31. Rousset, F. (2008). genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour* 8, 103-106.
32. Saito, Y., Shiraishi, S., Tanimoto, T., Yin, L., and, S.W., and Ide, Y. (2002). Genetic diversity of *Populus euphratica* populations in northwestern China determined by RAPD DNA analysis. *New Forests* 23.
33. Schmidt, D.A., Campbell, N.R., Govindarajulu, P., Larsen, K.W., and Russello, M.A. (2020). Genotyping-in-Thousands by sequencing (GT-seq) panel development and application to minimally invasive DNA samples to support studies in molecular ecology. *Mol Ecol Resour* 20, 114-124.
34. Shen, D., Bo, W., Xu, F., and Wu, R. (2014). Genetic diversity and population structure of the Tibetan poplar (*Populus szechuanica* var. *tibetica*) along an altitude gradient. *BMC Genetics* 15, 1471-2156.
35. Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30, 2725-2729.
36. Verma, S., Gupta, S., Bandhiwal, N., Kumar, T., Bharadwaj, C., and Bhatia, S. (2015). High-density linkage map construction and mapping of seed trait QTLs in chickpea (*Cicer arietinum* L.) using Genotyping-by-Sequencing (GBS). *Sci Rep* 5, 17512.
37. Vonlanthen, Z.X., and Bruehlheide H (2010). Clonal structure and genetic diversity of three desert phreatophytes. *Am J Bot* 97, 234-242.
38. Wang, J., Kallman, T., Liu, J., Guo, Q., Wu, Y., Lin, K., and Lascoux, M. (2014).

Speciation of two desert poplar species triggered by Pleistocene climatic oscillations. *Heredity* (Edinb) **112**, 156-164.

39. Wang, J., Wu, Y., Ren, G., Guo, Q., Liu, J., and Lascoux, M. (2011). Genetic differentiation and delimitation between ecologically diverged *Populus euphratica* and *P. pruinosa*. *PLoS One* **6**, e26530.

40. Wang, Q., Abbott, R.J., Yu, Q.S., Lin, K., and Liu, J.Q. (2013). Pleistocene climate change and the origin of two desert plant species, *Pugionium cornutum* and *Pugionium dolabratum* (Brassicaceae), in northwest China. *New Phytol* **199**, 277-287.

41. Wang, S.J., Chen, B.H. and Li, H.Q. (1995). *Euphrates poplar forest*. China Environmental Science Press.

42. Wu, Y., Wang, J., and Liu, J. (2008). Development and characterization of microsatellite markers in *Populus euphratica* (Populaceae). *Mol Ecol Resour* **8**, 1142-1144.

43. Xu, F., Feng, S., Wu, R., and Du, F.K. (2013). Two highly validated SSR multiplexes (8-plex) for Euphrates' poplar, *Populus euphratica* (Salicaceae). *Mol Ecol Resour* **13**, 144-153.

44. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82.

45. Ye, J.W., Zhang, Z.K., Wang, H.F., Bao, L., and Ge, J.P. (2019). Phylogeography of *Schisandra chinensis* (Magnoliaceae) Reveal Multiple Refugia With Ample Gene Flow in Northeast China. *Front Plant Sci* **10**, 199.

46. Zeng, Y.F., Zhang, J.G., Abuduhamiti, B., Wang, W.T., and Jia, Z.Q. (2018). Phylogeographic patterns of the desert poplar in Northwest China shaped by both geology and climatic oscillations. *BMC Evol Biol* **18**, 75.

47. Zhang, H.X., Li, X.S., Wang, J.C., and Zhang, D.Y. (2020). Insights into the aridification history of Central Asian Mountains and international conservation strategy from the endangered wild apple tree. *Journal of Biogeography* **48**, 332-344.

48. Zheng, Y., Jiao, P., Zhao, Z., and Li, Z. (2016). Clonal growth of *Populus Pruinosa* Schrenk and its role in the regeneration of riparian forests. *Ecological Engineering* **94**, 380-392.

FIGURE LEGENDS

Figure 1. Geographical distributions of the natural populations of *P. euphratica* and *P. pruinosa* analyzed in this study. Different font colors represent the two species: black for *P. euphratica* populations, light blue for *P. pruinosa* populations, and red for overlapping populations. The map was created using the ArcMap package in ArcGIS ver. 10.8 (<http://www.esri.com/software/arcgis>). The populations are described in Table S1.

Figure 2. Population stratification for $K = 2, 3$, and 4 determined using STRUCTURE software.

Figure 3. PCA based on genetic distance showing four clustered subpopulations within *P. euphratica* and *P. pruinosa*.

Figure 4. Neighbor-joining phylogenetic tree based on a genetic distance matrix representing the grouping of 1,183 *P. euphratica* and 437 *P. pruinosa* individuals.

Figure 5. Tests of isolation-by-distance and isolation-by-environment among 58 *P. euphratica* populations (A and B) and 24 *P. pruinosa* populations (C and D), respectively. Correlation coefficients were analyzed with data matrixes of genetic distance $\times$ geographical distance and genetic distance $\times$ environmental distance from pairwise populations.

Figure 6. Effects of geographical and environmental factors on genetic diversity in *Populus* populations. (A–C) *P. euphratica*. (D–F) *P. pruinosa*.

Table 1. Comparison of genetic diversity among and within species.

ID	Pop	H_e	H_o	I	PIC
GNM	Pe	0.212	0.210	0.323	0.171
NX	Pe	0.206	0.197	0.317	0.167
QH	Pe	0.112	0.205	0.161	0.086
SX	Pe	0.216	0.210	0.339	0.176
	Pp	0.256	0.194	0.399	0.210

Pe, *P. euphratica*; Pp, *P. pruinosa*; GNM, Gansu, Ningxia, and Inner Mongolia; NX, northern Xinjiang; QH, Qinghai; SX, southern Xinjiang.

Table 2. Comparison of genetic differentiation among and within species.

ID	Pop	Pp	NX	SX	QH	M	GN
	Pp	0.000					
	Pe						
NX	Pe	0.160	0.000				
SX	Pe	0.146	0.044	0.0			
QH	Pe	0.118	0.042	0.0	0.000		
	Pe			0.0			
GNM	Pe	0.131	0.022	0.10	0.064		0.000

Pe, *P. euphratica*; Pp, *P. pruinosa*; GNM, Gansu, Ningxia, and Inner Mongolia; NX, northern Xinjiang; QH, Qinghai; SX, southern Xinjiang.

Table 3. Analysis of molecular variance of the genetic variation among and within *P. euphratica* populations.

Source	df	SS	MS	Var.	Est.	%
Among populations	4	92843.949	2321	0.987	40.	1
Among individuals	1574	526277.637	334.	518	99.	3
Within individuals	1579	212664.000	134.	837	134	4
Total	3157	831785.586	683	.683	275	1
				.038	00%	
F_{ST}	0.147					
N_m	1.447					

Supplementary file

Supplementary Table S1. Overview of information about the samples used in this study.

Supplementary Table S2. The primers used for each target SNP site.

Supplementary Table S3. Spearman's rank correlation coefficients (r) for the seven remaining bioclimatic variables in *P. euphratica*.

Supplementary Table S4. Spearman's rank correlation coefficients (r) for the seven remaining bioclimatic variables in *P. pruinosa*.

Supplementary Table S5. Information about the SX1 samples analyzed by PCA.

Supplementary Table S6. Genetic diversity analysis of *P. euphratica* samples from 59 natural populations.

Supplementary Table S7. Genetic diversity analysis of *P. pruinosa* samples from 26 natural populations.

Supplementary Table S8. Genetic differentiation coefficients among *P. euphratica* populations.

Supplementary Table S9. Genetic differentiation coefficients among *P. pruinosa* populations.

Supplementary Figure S1. Distribution of selected SNPs on the 19 *P. euphratica* chromosomes.

Supplementary Figure S2. Phenotypic variation in the leaves of *P. euphratica* and *P. pruinosa* natural populations in Maigaiti County, southern Xinjiang.