

# Genetic diversity and population structure of black cottonwood (*Populus deltoides*) revealed using simple sequence repeat markers

**Cun Chen**

Chinese Academy of Forestry Research Institute of Forestry

**Yanguang Chu**

Chinese Academy of Forestry Research Institute of Forestry

**Changjun Ding**

Chinese Academy of Forestry Research Institute of Forestry

**Xiaohua Su** (✉ [suxh@caf.ac.cn](mailto:suxh@caf.ac.cn))

- <https://orcid.org/0000-0002-1497-854X>

**Qinjun Huang**

Chinese Academy of Forestry Research Institute of Forestry

---

## Research article

**Keywords:** *Populus deltoides*, genetic diversity, population structure, simple sequence repeat

**Posted Date:** December 20th, 2019

**DOI:** <https://doi.org/10.21203/rs.2.10562/v3>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on January 6th, 2020. See the published version at <https://doi.org/10.1186/s12863-019-0805-1>.

# Abstract

**Background:** Black cottonwood (*Populus deltoides*) is one of the keystone forest tree species, and has become the main breeding parents in poplar hybrid breeding. However, the genetic diversity and population structure of the introduced resources are not fully understood.

**Results:** In the present study, five loci containing null alleles were excluded and 15 pairs of SSR (simple sequence repeat) primers were used to analyze the genetic diversity and population structure of 384 individuals from six provenances (Missouri, Iowa, Washington, Louisiana, and Tennessee (USA), and Quebec in Canada) of *P. deltoides*. Ultimately, 108 alleles ( $N_a$ ) were detected; the expected heterozygosity ( $H_e$ ) per locus ranged from 0.070 to 0.905, and the average polymorphic information content (PIC) was 0.535. The provenance 'Was' had a relatively low genetic diversity, while 'Que', 'Lou', and 'Ten' provenances had high genetic diversity, with Shannon's information index ( $I$ ) above 1.0. The mean coefficient of genetic differentiation ( $F_{st}$ ) and gene flow ( $N_m$ ) were 0.129 and 1.931, respectively. Analysis of molecular variance (AMOVA) showed that 84.88% of the genetic variation originated from individuals. Based on principal coordinate analysis (PCoA) and STRUCTURE cluster analysis, individuals distributed in the Mississippi River Basin were roughly classified as one group, while those distributed in the St. Lawrence River Basin and Columbia River Basin were classified as another group. The cluster analysis based on the population level showed that provenance 'low' had a small gene flow and high degree of genetic differentiation compared with the other provenances, and was classified into one group. There was a significant relationship between genetic distance and geographical distance.

**Conclusions:** *P. deltoides* resources have high genetic diversity and there is a moderate level of genetic differentiation among provenances. Geographical isolation and natural conditions may be the main factors causing genetic differences among individuals. Individuals reflecting population genetic information can be selected to build a core germplasm bank. Meanwhile, the results could provide theoretical support for the scientific management and efficient utilization of *P. deltoides* genetic resources, and promote the development of molecular marker-assisted breeding of poplar.

## Background

Genetic diversity is an important component of biodiversity, and is the basis of ecosystem and species diversity [1-3]. Understanding the genetic diversity and structure of germplasm resources aids efficient and rational development, and the utilization of germplasm resources on the premise of effective protection [4-7]. The growth characteristics of trees involve reaching a large size with a long growth cycle; therefore, it is necessary to analyze their genetic diversity and population structure. To fully understand the genetic information of germplasm resources, screening representative individuals and constructing a core germplasm bank for protection and utilization can shorten the breeding process and accelerate genetic improvement [8]. Currently, researchers regard the study of forest genetic diversity and population structure as important basic research, and such studies have been carried out on a variety of species. Among forest trees, the genome of *Populus trichocarpa* was the first to be sequenced [9]; therefore, there

have been relatively more studies on its genetic information, including its genetic diversity and population structure [10,11]. In addition, similar studies have been carried out on other tree species, such as *P. nigra* L. [12-14], *P. simonii* [15], *P. tremuloides* [16], *P. balsamifera* [17], *P. cathayana* Rehd [18], *P. euphratica* [19], *P. tomentosa* [20] and *P. szechuanica* var. *tibetica* [21].

Many methods and techniques have been developed to advance population genetic diversity research, among which molecular tools play an important role in the management and utilization of genetic resources. In particular, simple sequence repeat (SSR) molecular marker technology is an ideal method because of its simple operation, co-dominance, high resolution, polymorphism, and repeatability [22]. Many species-specific SSR markers have been developed for *P. simonii* [23, 24] and *P. tomentosa* [25-27] by analyzing functional gene sequences, and these markers can be applied to the study of other poplar species.

*P. deltoides* is widely distributed from the Mississippi River to southern Canada in North America, and is one of the keystone forest tree species with important ecological value. This species has been widely used in forest breeding research and its genetic resources have become the main gene donors of poplar cultivars. These cultivars are mainly used to provide feedstocks for pulp, fiber, and bioenergy industries [28]. Furthermore, this tree is currently one of the most suitable for short rotation industrial timber intensive management of woody crops in the mid-latitude areas of the world. Our team collected *P. deltoides* germplasm resources through international exchange and cooperation in 2009, and a germplasm bank of *P. deltoides* was established in China. Production practice showed that the black poplar has a dominant position in the poplar plantation; thus, *P. deltoides* and its hybrid with *P. nigra* L. have become the main commercial poplar trees in China.

Little research has been performed on the genetic diversity and population structure of *P. deltoides*. Fahrenkrog *et al.* [29] reported the first population genomics study for *P. deltoides* distributed in the Mississippi River Basin to determine its genetic diversity and adaptation potential. In the present study, we analyzed 384 *P. deltoides* individuals preserved in the germplasm bank, which were originally distributed in 27 collection sites for 6 provenances. SSR molecular marker technology was used to study their genetic diversity and population structure. Some of the selected primers were developed based on functional genes sequences, which could be used to study gene function of *P. deltoides*. In this way, more information could be provided for marker-assisted breeding in combination with phenomics. Meanwhile, other (universal) primers were obtained from the International Populus Genome Consortium (IPGC, [http://www.ornl.gov/sci/ipgc/ssr\\_resource.htm](http://www.ornl.gov/sci/ipgc/ssr_resource.htm)) and Washington University (Poplar Molecular Genetics Cooperative, <http://poplar2.cfr.washington.edu>). Using this strategy, the genetic diversity of the population could be reflected objectively and comprehensively. In addition, based on the results, a core germplasm bank of *P. deltoides* could be built, allowing *P. deltoides* resources to be protected, managed, and utilized more scientifically and rationally.

## Results

## Polymorphic SSR primers

In this study, 20 polymorphic SSR primer pairs were preliminary selected from 145 poplar SSR primer pairs for further analysis (Additional file 1: Table S1). Using two methods to detect the null alleles of the loci (Additional file 1: Table S2) [30, 31], we found that the null alleles of five loci (SSR15, SSR42, SSR54, SSR65, and SSR76) could affect the analysis results; therefore, the data of these five loci were excluded from the subsequent analysis. Finally, 15 SSR primer pairs were selected to analyze the genetic diversity and population structure of *P. deltoides*.

## Microsatellite polymorphisms

A total of 108 alleles were detected using 15 SSR primer pairs among 384 *P. deltoides* individuals. The number of alleles ( $N_a$ ) per locus ranged from 2 (SSR85) to 19 (SSR126), with a mean of 7.2. The average effective number of alleles ( $N_e$ ) per locus was 3.48, ranging from 1.08 at SSR85 to 10.40 at SSR126. Locus SSR126 provided relatively large amounts of genetic information and its Shannon's information index ( $I$ ) was 2.533. The means of the observed heterozygosity ( $H_o$ ) and expected heterozygosity ( $H_e$ ) were 0.509 and 0.579, respectively. In addition, except for locus SSR104, the false-homozygous phenomenon existed in the other loci ( $H_o > H_e$ ). The polymorphic information content ( $PIC$ ), as one measurement of genetic diversity, was between 0.068 (SSR85) and 0.896 (SSR120), with an average of 0.535. Six loci showed significant deviations from the Hardy-Weinberg equilibrium (HWE), which indicated that there was genetic differentiation among the provenances (Additional file 1: Table S3).

The heat map of polymorphic SSR loci revealed the genetic diversity information of the loci, with the color richness being related to the degree of polymorphism. Analysis of the results (Fig. 1) showed that loci SSR85, SSR104, and SR143 had monotone colors, indicating relatively poor levels of polymorphism, whereas other loci were rich in polymorphisms. In addition, homozygous and heterozygous information for the loci were also clearly and intuitively expressed in the heat map.

## Population genetic diversity

Based on the result of the population genetic diversity analysis (Additional file 1: Table S4), we found that the 'Was' population had relatively low genetic diversity ( $N_a = 2.60$ ,  $N_e = 1.96$ ,  $I = 0.617$ ,  $H_o = 0.365$ ,  $H_e = 0.358$ ). The  $I$  values of the 'Que', 'Lou', and 'Ten' populations were 1.032, 1.131, and 1.175, respectively, and their genetic diversities were relatively high. Meanwhile these three populations had private alleles. The HWE results showed that the 'Que' and 'Ten' populations deviated significantly from the equilibrium.

The abundance of amplified fragments and the differences in the amplified fragments at each point in each population could be displayed visually using a heat map of the population genetic diversity (Fig. 2), which reflected the genetic variation of the population. The individuals from the 'Was' population were homozygous at the SSR58, SSR80, SSR85, SSR104, SSR105, and SSR143 loci, and its genetic diversity was poor. The size and proportion of amplified fragments could reveal the main genotypes and their differentiation in each population at each locus. Taking locus SSR143 as an example, single amplified

fragments were obtained from the 'Mis', 'low', and 'Was' populations; some individuals in the 'Que' and 'Ten' populations were deleted on the basis of the main amplified fragments; and some individuals in the 'Lou' population produced amplicons with increased length.

### Population genetic differentiation and genetic variation

The *P. deltooides* population showed abundant genetic diversity and genetic differentiation. The within-population ( $F_{is}$ ) and inter-population ( $F_{it}$ ) inbreeding coefficients were used as indicators to evaluate the degree of population neatness. The inter-population genetic fraction coefficient ( $F_{st}$ ) was used as an indicator to evaluate the level of genetic differentiation of the populations. The existence of large gene flow ( $N_m > 1$ ) among populations weakened the possibility of genetic drift, which would decrease the degree of genetic differentiation among populations. By contrast, when  $N_m < 1$ , the genetic differentiation among populations increases [32]. The average  $F_{is}$  and  $F_{it}$  values of *P. deltooides* were 0.058 and 0.182, respectively, indicating that there was a loss of heterozygosity in the population, and there was inbreeding among the populations. The  $F_{st}$  between populations ranged from 0.062 to 0.205, with an average of 0.129, indicating that there was moderate genetic differentiation among the populations (Table 1). Meanwhile, higher gene flow (mean = 1.931, Table 1) among populations prevented genetic differentiation among populations to a certain extent.

The result of analysis of molecular variance (AMOVA) showed that 84.88% of the total genetic variation originated from individuals, while 11.49% came from populations, and only 3.64% was ascribed to differences among individuals within the populations (Table 2). The results were consistent with the previous analysis results, which suggested that the genetic diversity of *P. deltooides* was mainly caused by genetic differences among individuals.

Analysis using the Mantel test showed that there was a very significant relationship between genetic distance and geographical distance between points ( $R^2 = 0.1135$ ,  $P = 0.001$ , Fig. 3). Analysis of the correlation between the genetic distance and geographical distance of latitude, showed an extremely significant correlation and the fitting degree was relatively high ( $R^2 = 0.1867$ ,  $P = 0.001$ , Fig. 3).

### Genetic structure

The results of principal coordinate analysis (PCoA) showed that the 384 individuals of *P. deltooides* could be divided into two groups (Fig. 4a). Group I mainly included the individuals from 'Que' and 'Was', and a few individuals from the 'low' provenance. Group II mainly included the individuals of the 'Ten', 'Lou', 'Mis', and 'low' provenances. The individuals from the 'Que' provenance were widely distributed, which indicated that 'Que' had abundant genetic diversity. STRUCTURE cluster analysis showed that when  $K = 2$ , the  $\Delta K$  value was relatively large (Fig. 4b), which indicated that the 384 *P. deltooides* individuals could be divided into two different groups. The genetic structure map (Fig. 4c) showed that most individuals from 'Que' and 'Was' belonged to the red group (Group I), while most individuals from 'Ten', 'Lou', and 'low' belonged to the green group (Group II). Based on the analysis of the matrix of its estimated membership probability

(Q-matrix) when  $K = 2$ , we could determine the composition of the red and green groups (Additional file 1: Table S5).

To understand the genetic structure among the six populations of *P. deltooides*, we carried out PCoA and cluster analysis between the populations, based on the unweighted pair-group method with arithmetic means (UPGMA) on the provenance populations (Fig. 5a, b). In addition, an unrooted tree was drawn (Fig. 5c). The results showed that the 'Lou' and 'Ten' provenances were clustered together; the 'Mis', 'Was,' and 'Que' provenances were clustered together; and 'low' was independent of the other provenance populations.

## Discussion

### Null alleles

SSR molecular markers have been widely used to study population genetic diversity and genetic structure. However, the existence of null alleles might reduce the population genetic diversity and increase genetic differentiation among populations, which will have a significant impact on the results of the study [33]. In this study, five pairs of primers with a high frequency of null alleles were identified. To clearly understand the effect of loci containing null alleles on the population genetic analysis results, we compared the changes of major genetic parameters before and after deletion of these loci. We found the difference between the  $H_o$  and  $H_e$  values increased, the  $F_{st}$  increased, and the  $N_m$  decreased in the presence of the null allele loci (Table 3). These changes indicated that the five loci had null alleles that likely contributed to the positive genetic fraction coefficient, and affected the interpretation of the results [34].

### SSR Primer Screening

The SSR primers based on functional gene sequence analysis of poplar have different characteristics in different varieties. In this study the amplification motifs of two primer pairs, SSR15 and SSR42, showed some differences. The SSR15 locus was located in the promoter region of the *PsHsf16* gene of the *P. simonii* Hsf family. In a previous study [23], its repeat unit was ATTT; however, in this study, the repeat motif was ATT (Fig. 6a). The SSR42 locus was in the intron region of the *PtCesA6* gene in *P. tomentosa*, and its repeating motif was TTCTCC [25], whereas, in this study, the amplifying motif was TC (Fig. 6b). Meanwhile, we observed that the molecular markers of the functional genes of other poplar varieties screened in this study were likely to generate null alleles. These results indicated that the marker characteristics of the same primers might change when they are applied to different experimental materials [35]. Therefore, the stability and polymorphism of primers should be further verified when screening primers that have been developed based on the analysis of functional gene sequences. Of course, it would be better to develop species-specific SSR markers, which is conducive to further research.

### Genetic diversity and variation

A better understanding of the genetic diversity, genetic variation, and genetic regulation of a population is essential for its proper management and conservation, especially in the face of current climate change, and the genetic evaluation of groups such as forests is particularly important [36]. Fahrenkrog *et al.* performed the first population genomics study for 425 unrelated individuals of *P. deltooides* distributed in 13 states of the southeastern United States. By assessing population structure, population differentiation, genetic diversity and adaptation in *P. deltooides*, they found that the differentiation between subpopulations of the natural *P. deltooides* population was weak ( $F_{st} = 0.022-0.106$ ); however, the genetic diversity was high [29]. Furthermore, genome-wide association studies (GWASs) were also used to dissect the genetic regulation of eight growth and wood composition traits in *P. deltooides*, and single-nucleotide polymorphisms were detected by targeted resequencing of 18,153 genes in a population of 391 unrelated individuals. The authors found that both common and low frequency variants need to be considered to understand the genetic regulation of complex traits [37]. In the present study, the genetic diversity and population structure of 384 individuals of *P. deltooides* from six provenances in the Saint Lawrence River Basin (Quebec, Canada), Columbia River Basin (Washington, USA) and Mississippi River Basin (Missouri, Iowa, Tennessee and Louisiana, USA) were analyzed using the SSR marker technique. The results showed that the *P. deltooides* had abundant genetic diversity and moderate genetic differentiation. In addition, the degree of genetic differentiation was relatively high, probably because the materials in this study were from three basins in the main distribution area of *P. deltooides*, and the materials in Fahrenkrog's study were mainly from the Mississippi River Basin [29, 37]. Our research also showed that geographical isolation hindered gene exchange among individuals from different provenances and aggravated genetic differentiation. In addition, the genetic distance between the provenances in the north ('Que', 'Was') and the south ('Ten', 'Lou') was relatively large, while the genetic distance between the provenances in the east ('Que') and the west ('Was') was relatively small. The reason for this result was consistent with the climate distribution in North America.

## Population structure

In this study, we analyzed the population structure of *P. deltooides* at different levels. At the individual level, PCoA and STRUCTURE cluster analysis were performed on 384 unrelated individuals, which divided them into two groups. Most of the individuals distributed in the Mississippi River Basin were grouped together (Group II, Fig. 4a, c), and individuals in different states of the Mississippi River Basin had high genetic similarity, which was consistent with the results of Fahrenkrog *et al.* [29, 37]. The genetic differentiation between different basins indicated that geographical isolation restricts gene exchange among populations [38]. Individuals distributed in the Saint Lawrence River Basin and the Columbia River Basin were grouped into the other group (Group I, Fig. 4a, c). These two locations might have the same type of genetic variation because of their similar environment and latitudes. Taking the provenances as the research object, the 'Lou' and 'Ten' provenances were clustered into one group, the 'Mis', 'Was' and 'Que' provenances were clustered into another group, and the 'low' provenance formed a third group (Fig. 5). To determine the reasons for the differences between the results of individuals and provenances analysis, we analyzed  $N_m$  and  $F_{st}$  among the six provenances (Table 4). We observed that the  $N_m$  between 'low'

and the others was relatively small and the  $F_{st}$  was large, which could explain why the 'low' provenance was separately classified into one group [39]. Eighteen individuals from 'low' were selected, and PCoA of 384 individuals of *P. deltooides* showed that they were distributed widely in the dimension of the first principal component, and most individuals were on the edge of Group II (Fig. 4a). We hypothesized that individuals from 'low' might have been subjected to some external selection pressure, which has resulted in genetic variation in a certain direction. This selection pressure also exists in other provenances; however, the selection pressure is relatively weak and did not lead to directional mutations of genes. Therefore, when analyzing population structure, we should not neglect the influence of external environmental conditions on the structure. At the same time, geographical isolation has restricted gene exchange between 'low' and other provenances.

## Management and Utilization of Germplasm Resources

The management of germplasm resources is a complex task. We should have a full understanding of their genetic information, morphological diversity, and adaptability. Meanwhile, germplasm resources need to be effectively identified to prevent redundancy between resources [40, 41]. Germplasm banks play an important role in the conservation, management, and utilization of germplasm resources that is critical for the development of plant breeding [42]. Molecular genetic markers are widely used in germplasm identification, and they are important in the construction and management of germplasm banks [43]. For example, the single nucleotide polymorphism (SNP) markers were used to identify *Dimocarpus longan* L. germplasm [44] and *Discorea alata* L. germplasm [45]. Reyes-Valdés *et al.* offered an integrated view of accession rarity and allele specificity in germplasm banks for management and conservation [43]. Storme *et al.* [46] analyzed 675 *P. nigra* L. accessions from nine European gene banks using SSR markers, amplified fragment length polymorphism (AFLP) markers and isozyme systems to estimate the extent of duplication and the genetic diversity within and between banks. To better manage and utilize *P. tomentosa* [20] and *P. simonii* [47] resources, their genetic diversity had been studied using SSR markers. Combined with the analysis of phenotypic diversity of *P. deltooides*, we may identify the key polymorphic loci associated with the traits through correlation analysis, which is conducive to the development of molecular marker-assisted breeding or detection of target genes in the near future. In addition, a core germplasm bank could be built to rationally manage, preserve, and utilize the *P. deltooides* resources.

## Conclusions

In the present study, the genetic diversity and population structure of *P. deltooides* germplasm resources were analyzed. The results showed that they had abundant genetic diversity. However, there were also some differences in genetic diversity among different provenances, with moderate genetic differentiation (mean  $F_{st}$  = 0.129, Table 1). The genetic variation mainly came from individuals in different provenances. Geographical isolation was the main reason for the differences among the provenances. Based on the results, representative individuals could be selected to form a core germplasm bank of *P. deltooides* to improve the selection efficiency of hybrid parents of poplar and to lay a scientific foundation for the conservation and breeding of poplar germplasm resources.

# Methods

## Plant materials and DNA extraction

Cuttings of *P. deltoides* were collected from the germplasm bank (35°55'39"N, 116°53'59"E) that was established through international exchange and cooperation of germplasm resources in 2009, among which the resources came from its main distribution areas. The materials have been formally identified by the State Forestry and Grassland Administration, People's Republic of China, under the identification number 2009-56. In this study, 384 unrelated individuals were randomly selected from six provenances. When collecting the materials, several collection sites were selected in each provenance, and then individual materials were collected near the collection sites. The geographical distance between individuals was more than 100 meters. Six individuals were selected from two sampling sites in Missouri ('Mis'). Eight sampling sites were selected in Louisiana ('Lou') and five in Tennessee ('Ten'); 108 individuals and 75 individuals were selected, respectively. Eighteen unrelated individuals were from Iowa ('Iow'). The four provenances were located in Mississippi River basin. Seventeen unrelated individuals were from Washington ('Was'), which is located in the Columbia River basin. In addition, 160 unrelated individuals were from ten sampling sites in the Saint Lawrence River basin (Quebec, Canada, 'Que') (Fig. 7, Additional file 1: Table S6).

The plant materials were cultured in the greenhouse of the Chinese Academy of Forestry (40°0'10"N 116°14'38"E), and the leaves were collected during the growing season and stored in a refrigerator at -40 °C for DNA extraction. Total genomic DNA was extracted according to the highly efficient modified cetyltrimethylammonium bromide (CTAB) method [48]. The quality and integrity of the extracted genomic DNA were assessed using 1% agarose gel electrophoresis. The DNA concentration was then determined using a NanoDrop-2000 ultramicro-spectrophotometer (Thermo Fisher Scientific Waltham, MA, USA). Finally, the DNA was diluted to 50 ng· $\mu\text{L}^{-1}$  and stored at -20 °C for polymerase chain reaction (PCR) amplification.

## Primer selection and PCR amplification

In this study, DNA of 20 unrelated individuals *P. deltoides* were selected randomly for SSR primer screening, and 145 pairs of primers were used for screening. Seventy-five of the primer pairs were developed by researchers based on the analysis of functional gene sequences in poplar [23-27, 49]. Another 70 pairs of primers were obtained from the International Populus Genome Consortium (IPGC, [http://www.ornl.gov/sci/ipgc/ssr\\_resource.htm](http://www.ornl.gov/sci/ipgc/ssr_resource.htm)) and Washington University (Poplar Molecular Genetics Cooperative, <http://poplar2.cfr.washington.edu>), which have used by other researchers in related studies [47, 50-52]. The stability and polymorphism of the SSR primers were preliminarily screened by PCR amplification, 2% agarose gel electrophoresis, and 8% non-denaturing polyacrylamide gel electrophoresis.

The PCR amplification reaction system for all SSR markers comprised a 25  $\mu\text{L}$  mixed system containing 2.5  $\mu\text{L}$  of 10 × buffer ( $\text{Mg}^{2+}$  plus), 1.8  $\mu\text{L}$  of dNTP mixture, 1  $\mu\text{L}$  of forward primer (10  $\mu\text{mol}\cdot\text{L}^{-1}$ ), 1  $\mu\text{L}$  of reverse primer (10  $\mu\text{mol}\cdot\text{L}^{-1}$ ), 0.25  $\mu\text{L}$  of *Taq* polymerase (5 U· $\mu\text{L}^{-1}$ ), 1  $\mu\text{L}$  of DNA template (50 ng· $\mu\text{L}^{-1}$ ), and

17.45  $\mu\text{L}$  of ddH<sub>2</sub>O. The PCR amplification procedure comprised 94 °C for 3 minutes; 35 cycles of 94 °C for 30 seconds, annealing at the annealing temperature of each primer pair for 30 seconds, elongation at 72 °C for 45 seconds; and a final extension at 72 °C for 10 minutes.

The forward primers for the preliminarily screened polymorphic loci were labeled with fluorescent dyes (5-HEX or 5-FAM) and their PCR products were separated by capillary electrophoresis using an ABI 3730xl DNA analyzer (Applied Biosystems, Foster City, CA, USA). A peak size map of the amplified fragments was obtained.

## Data analysis

The Gene-Marker 2.2.0 software (SoftGenetics LLC, USA) was used to read the peak maps of the amplified fragments of each polymorphic locus. The null allele frequencies of each locus were detected using Cervus 3.0.7 [30] and Microchecker 2.2.3 [31] software, and the common detection results were taken. The Cervus 3.0.7 software was also used to calculate the *PIC* of each locus.

The GeneAEx 6.503 [53] software was used to convert various file formats for different analysis and to calculate genetic diversity parameters, including the number of alleles ( $N_a$ ), the effective number of alleles ( $N_e$ ), Shannon's information index ( $I$ ), observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ), the number of private alleles, gene flow ( $N_m$ ), and the *F*-Statistics ( $F_{is}$ ,  $F_{it}$ ,  $F_{st}$ ). The HWE tests across all provenances were performed using Genepop 4.7.0 [54]. The heat maps of each locus were drawn using OmicShare tools ([www.omicshare.com/tools](http://www.omicshare.com/tools)). With the help of the Arlequin 3.5.2.2 [55] and GeneAEx 6.503 software, AMOVA was carried out to partition the genetic variances into three levels: Among populations, among individuals within populations, and within individuals.

We further calculated the genetic distance between individuals and populations using the GeneAEx 6.503 software for PCoA. A cluster analysis between populations, based on UPGMA, was also developed using the NTSYS-pc 2.10e [56] software. An unrooted tree was constructed based on pairwise standard genetic distances [57], using the least squares algorithm with 10,000 bootstrap replicates, and these processes were generated and analyzed using PHYLIP 3.6 [58] software.

The population genetic structure was analyzed using STRUCTURE 2.3.4 [59] software, using a model-based clustering algorithm that implements a Bayesian framework and the Markov chain Monte Carlo (MCMC) algorithm. To confirm the optimum number of subpopulations ( $K$ ), 10 independent runs for each value of  $K$ , ranging from 2 to 5, were conducted. Each run consisted of a burn-in period of 100,000 steps followed by 1,000,000 MCMC iterations. The  $\Delta K$  parameter, which was based on the rate of change in the log probability of data between successive  $K$  values, was estimated to determine the best  $K$ , based on the model developed by Evanno *et al.* [60].

The Mantel test was performed using the GeneAEx 6.503 software to analyze the correlation between Nei's genetic distance and geographical distance. The geographical distance between different sites was

calculated according to the latitude and longitude using Vincenty's formula ([www.movable-type.co.uk/scripts/latlong-vincenty.html](http://www.movable-type.co.uk/scripts/latlong-vincenty.html)).

## Abbreviations

SSR: simple sequence repeat; *PIC*: polymorphic information content; AMOVA: analysis of molecular variance; PCoA: principal coordinate analysis; 'Mis': the provenance population in Missouri, USA; 'low': the provenance population in Iowa, USA; 'Was': the provenance population in Washington State, USA; 'Que': the provenance population in Quebec, Canada; 'Lou': the provenance population in Louisiana, USA; 'Ten': the provenance population in Tennessee, USA; HWE: Hardy-Weinberg equilibrium;  $F_{IS}$ : inbreeding coefficient within-population;  $F_{IT}$ : inbreeding coefficient inter-population;  $F_{ST}$ : inter-population genetic fraction coefficient;  $N_m$ : gene flow; UPGMA: unweighted pair-group method with arithmetic means.

## Declarations

### Ethics approval and consent to participate

The experimental materials used in this study have been formally identified by the State Forestry and Grassland Administration, People's Republic of China, under the identification number 2009-56.

### Consent for publication

Not applicable.

### Availability of data and material

All data generated or analyzed during this study are included in this published article [and its additional files]. Genetic information of polymorphic loci has been uploaded as Additional file 1: Table S7.

### Competing interests

The authors declare that they have no competing interests.

### Funding

This study was supported by the Basic Research Fund of RIF (Grant No. CAFYBB2018SY007). The funder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Authors' contributions

QH and XS conceived and designed the experiments. CC and CD collected the experimental materials. YC assisted in screening primers. CC performed the experiments, analyzed the data, and was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

Not applicable.

## References

- [1] Zhang C, Vornam B, Volmer K, *et al.* Genetic diversity in aspen and its relation to arthropod abundance. *Frontiers in Plant Science*. 2015;5:806.
- [2] Haddad NM, Crutsinger GM, Gross K, *et al.* Plant diversity and the stability of foodwebs. *Ecology Letters*. 2011;14(1):42-6.
- [3] Bailey JK, Schweitzer JA, Ubeda F, *et al.* From genes to ecosystems: a synthesis of the effects of plant genetic factors across levels of organization. *Philosophical Transactions of the Royal Society of London*. 2009;364(1523):1607-16.
- [4] Frankham R, Ballou JD, Briscoe DA, *et al.* Introduction to conservation genetics: glossary. *Zoologica Africana*. 2002;38(1):192.
- [5] Millar CI, Westfall RD. Allozyme markers in forest genetic conservation. *New Forests*. 1992;6(1-4):347-71.
- [6] Wuyun TN, Amo H, Xu JS, *et al.* Population Structure of and Conservation Strategies for Wild *Pyrus ussuriensis* Maxim. in China. *PLoS One*. 2015;10(8):e0133686.
- [7] Sork VL, Aitken SN, Dyer RJ, *et al.* Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes*. 2013;9(4):901-11.
- [8] Frankel OH, Brown AHD. Plant genetic resources today: a critical appraisal. In: Holden JHW, Williams JT, editors. *Crop Genetic Resources: Conservation and Evaluation*. London: Academic; 1984. p. 249-57.
- [9] Tuskan GA, DiFazio S, Jansson S, *et al.* The genome of black cottonwood, *Populus trichocarpa* (Torr & Gray). *Science*. 2006;313(5793):1596-604.
- [10] Evans LM, Slavov GT, Rodgers-Melnick E, *et al.* Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nature Genetics*. 2014;46(10):1089-96.
- [11] Holliday JA, Zhou L, Bawa R, *et al.* Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in *Populus trichocarpa*. *New Phytologist*. 2016;209(3):1240-51.
- [12] Guet J, Fabbri F, Fichot R, *et al.* Genetic variation for leaf morphology, leaf structure and leaf carbon isotope discrimination in European populations of black poplar (*Populus nigra* L.). *Tree physiology*.

2015;35(8):850-63.

[13] Arens P, Coops H, Jansen J, *et al.* Molecular genetic analysis of black poplar (*Populus nigra* L) along Dutch rivers. *Molecular Ecology*. 2010;7(1):11-8.

[14] Zhang YX, Zhang XR, Che Z, *et al.* Genetic diversity assessment of sesame core collection in China by phenotype and molecular markers and extraction of a mini-core collection. *BMC Genetics*. 2012;13(1):102.

[15] Wei ZZ, Du QZ, Zhang JF, *et al.* Genetic diversity and population structure in Chinese Indigenous Poplar (*Populus simonii*) populations using microsatellite markers. *Plant Molecular Biology Reporter*. 2013; 31(3):620-32.

[16] Callahan CM, Rowe CA, Ryel RJ, *et al.* Continental-scale assessment of genetic diversity and population structure in quaking aspen (*Populus tremuloides*). *Journal of Biogeography*. 2013;40(9):1780-91.

[17] Meirmans PG, Godbout J, Lamothe M, *et al.* History rather than hybridization determines population structure and adaptation in *Populus balsamifera*. *Journal of Evolutionary Biology*. 2017;30(11):2044-58.

[18] Lu ZX, Wang YH, Peng YH, *et al.* Genetic diversity of *Populus cathayana* Rehd populations in southwestern China revealed by ISSR markers. *Plant Science*. 2006;170(2):407-12.

[19] Wang J, Li Z, Guo Q, *et al.* Genetic variation within and between populations of a desert poplar (*Populus euphratica*) revealed by SSR markers. *Annals of Forest Science*. 2011;68(6):1143-9.

[20] Du QZ, Wang BW, Wei ZZ, *et al.* Genetic diversity and population structure of Chinese White Poplar (*Populus tomentosa*) revealed by SSR markers. *Journal of Heredity*. 2012;103(6):853-62.

[21] Shen DF, Bo WH, Xu F, *et al.* Genetic diversity and population structure of the Tibetan poplar (*Populus szechuanica* var *tibetica*) along an altitude gradient. *BMC Genetics*. 2014;15(1 Supplement):S11.

[22] Ebrahimi A, Zarei A, McKenna JR, *et al.* Genetic diversity of Persian walnut (*Juglans regia*) in the cold-temperate zone of the United States and Europe. *Scientia Horticulturae*. 2017;220:36-41.

[23] Wang L, Wang BL, Wei ZZ, *et al.* Development of 35 microsatellite markers from heat stress transcription factors in *Populus simonii* (Salicaceae). *American Journal of Botany*. 2012;99(9):357-61.

[24] Wei ZZ, Du QZ, Guo Q, *et al.* *DREB* gene and its application in analyzing population structure in *Populus simonii*. *Bulletin of Botany*. 2010;45(5):556-65.

[25] Du QZ, Zhang DQ, Li BL. Development of 15 novel microsatellite markers from cellulose synthase genes in *Populus tomentosa* (Salicaceae). *American Journal of Botany*. 2012;99(2):46-8.

- [26] Du QZ, Wang BW, Wang BL, *et al.* Development and evaluation of simple sequence repeat (SSR) loci from functional genes involved in wood formation in *Populus tomentosa*. *Scientia Silvae Sinicae*. 2010;46(11):8-15.
- [27] Du QZ, Pan W, Xu BH, *et al.* Polymorphic simple sequence repeat (SSR) loci within cellulose synthase (*PtoCesA*) genes are associated with growth and wood properties in *Populus tomentosa*. *New Phytologist*. 2013;197(3):763-76.
- [28] Stanton BJ, Neale DB, Li S. *Populus* breeding: from the classical to the genomic approach In: Jansson S, Bhalerao RP, Groover AT, editors. *Genetics and genomics of Populus*. New York: Springer, 2010.p. 309-48.
- [29] Fahrenkrog AM, Neves LG, Resende Jr MF, *et al.* Population genomics of the eastern cottonwood (*Populus deltoides*). *Ecology and evolution*. 2017;7(22):9426-40.
- [30] Kalinowski ST, Taper ML, Marshall TC. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*. 2007;16:1099-06.
- [31] Oosterhout CV, Hutchinson WFD, Wills D, *et al.* Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, 2004;4(3):535-8.
- [32] Slatkin M. Gene flow and the geographic structure of natural populations. *Science*. 1987;236(4803):787-9
- [33] Hoffman JI, Amos W. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology*. 2010;14(2):599-612.
- [34] Guo Q, Li XY, Yang SH, *et al.* Evaluation of the genetic diversity and differentiation of Black Locust (*Robinia pseudoacacia* L) based on genomic and expressed sequence Tag-Simple Sequence Repeats. *International journal of molecular sciences*. 2018;19(9):2492.
- [35] Ellegren H, Galtier N. Determinants of genetic diversity. *Nature Reviews Genetics*. 2016;17:422-33.
- [36] Porth I, Klápště J, McKown AD, *et al.* Evolutionary quantitative genomics of *Populus trichocarpa*. *PloS one*. 2015;10(11):e0142864.
- [37] Fahrenkrog AM, Neves LG, Resende MF Jr, *et al.* Genome-wide association study reveals putative regulators of bioenergy traits in *Populus deltoides*. *New Phytologist*. 2016;213(2):799.
- [38] Geraldine A, Farzaneh N, Grassa CJ, *et al.* Landscape genomics of *Populus trichocarpa*: the role of hybridization, limited gene flow, and natural selection in shaping patterns of population structure. *Evolution*. 2014;68(11):3260-80.

- [39] Bothwell HM, Cushman SA, Woolbright SA, *et al.* Conserving threatened riparian ecosystems in the American West: Precipitation gradients and river networks drive genetic connectivity and diversity in a foundation riparian tree (*Populus angustifolia*). *Molecular Ecology*. 2017;26(19):5114-32.
- [40] Jarkko K, Barbara V, William D, *et al.* Utilization and transfer of forest genetic resources: A global review. *Forest Ecology and Management*. 2014;333(Sp. Iss. SI):22-34.
- [41] Belaj A, De La Rosa R, Lorite IJ, *et al.* Usefulness of a new large set of high throughput EST-SNP markers as a tool for olive germplasm collection management. *Frontiers in plant science*. 2018;9:1320.
- [42] Singh N, Wu S, Raupp WJ, *et al.* Efficient curation of genebanks using next generation sequencing reveals substantial duplication of germplasm accessions. *Scientific reports*. 2019;9(1):650.
- [43] Reyes-Valdés MH, Burgueño J, Singh S, *et al.* An informational view of accession rarity and allele specificity in germplasm banks for management and conservation. *PloS one*. 2018;13(2):e0193346.
- [44] Wang B, Tan HW, Fang W, *et al.* Developing single nucleotide polymorphism (SNP) markers from transcriptome sequences for identification of longan (*Dimocarpus longan*) germplasm. *Horticulture research*. 2015;2:14065.
- [45] Cormier F, Mournet P, Causse S, *et al.* Development of a cost-effective single nucleotide polymorphism genotyping array for management of greater yam germplasm collections. *Ecology and evolution*. 2019;9(10):5617-5636.
- [46] Storme V, Broeck AV, Ivens B, *et al.* Ex-situ conservation of Black poplar in Europe: genetic diversity in nine gene bank collections and their value for nature development. *Theoretical and Applied Genetics*. 2004;108(6):969-981.
- [47] Wei ZZ, Zhang GY, Du QZ, *et al.* Association mapping for morphological and physiological traits in *Populus simonii*. *BMC Genetics*. 2014;15(S1):S3.
- [48] Doyle. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull*. 1987;19:11-5.
- [49] Wang BL, Wang BW, Chen QQ, *et al.* Identification of SSR loci from transcription factor genes expressed under abiotic stresses in poplar. *Scientia Silvae Sinicae*. 2011;47(8):67-74.
- [50] Huang LJ, Su XH, Zhang XH, *et al.* SSR molecular markers related to wood density and fibre traits in poplar. *Journal of Genetics and Genomics*. 2004;31(3):299-304.
- [51] Zhang XH, Su XH, Huang QJ, *et al.* Comparison of genetic variations of black poplar (*Populus nigra* L.) gene resource by microsatellite markers. *Forest Research*. 2006;19(4):477-83.

- [52] Politov DV, Belokon MM, Belokon YS, *et al.* Application of microsatellite loci for molecular identification of elite genotypes, analysis of clonality, and genetic diversity in Aspen *Populus tremula* L (Salicaceae). *International Journal of Plant Genomics*. 2015;3:1-11.
- [53] Peakall R, Smouse PE. GENALEX 6: genetic analysis in Excel Population genetic software for teaching and research. *Molecular Ecology Notes*. 2006;6:288-95.
- [54] Raymond M, Rousset F. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Heredity*, 1995;86:248-9.
- [55] Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*. 2010;10:564-7.
- [56] Rohlf F J. NTSYS pc2.1: Numerical taxonomy and multivariate analysis system version 2.1. New York: Applied Biostatistics Inc. 2000.
- [57] Nei M. Genetic Distance between Populations. *American Naturalist*. 1972;106(949):283-92.
- [58] Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences. Seattle: University of Washington. 2005.
- [59] Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945-59.
- [60] Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*. 2005;14:2611-20.

## Tables

**Table 1. Genetic differentiation coefficients and gene flow of *P. deltoides* resources**

Locus	$F_{is}$	$F_{it}$	$F_{st}$	$N_m$
SSR6	0.085	0.142	0.062	3.789
SSR25	-0.035	0.177	0.205	0.968
SSR47	0.199	0.302	0.128	1.703
SSR58	-0.149	0.067	0.188	1.083
SSR80	0.111	0.218	0.120	1.825
SSR85	0.280	0.328	0.066	3.535
SSR104	-0.067	0.077	0.135	1.608
SSR105	0.151	0.228	0.090	2.513
SSR117	-0.033	0.126	0.154	1.377
SSR120	-0.100	-0.003	0.088	2.586
SSR126	-0.016	0.111	0.125	1.752
SSR129	0.137	0.265	0.148	1.444
SSR132	-0.025	0.175	0.195	1.033
SSR139	-0.003	0.108	0.111	2.005
SSR143	0.328	0.413	0.126	1.740
Mean	0.058	0.182	0.129	1.931

$F_{is}$ : inbreeding coefficient within-population;  $F_{it}$ : inbreeding coefficient inter-population;  $F_{st}$ : inter-population genetic fraction coefficient;  $N_m$ : gene flow.

Table 2. Analysis of molecular variance (AMOVA) using 384 individuals from six populations of *P. deltooides* resources

Source of variation	<i>df</i>	Sum of squares	Variance of components	Percentage variation (%)	<i>P</i> -Value
Among populations	5	300.216	0.516	11.49	<0.001
Among individuals within populations	378	1566.201	0.163	3.64	<0.001
Within individuals	384	1465.500	3.816	84.88	<0.001
Total	767	3331.917	4.496	100.00	

*df*: degrees of freedom.

Table 3. Statistical table of genetic diversity parameters with and without null allele loci

	$H_o$ (pop.)	$H_e$ (pop.)	$F_{st}$	$N_m$
Delete null allele loci	0.466	0.487	0.129	1.931
Include null allele loci	0.444	0.515	0.131	1.859

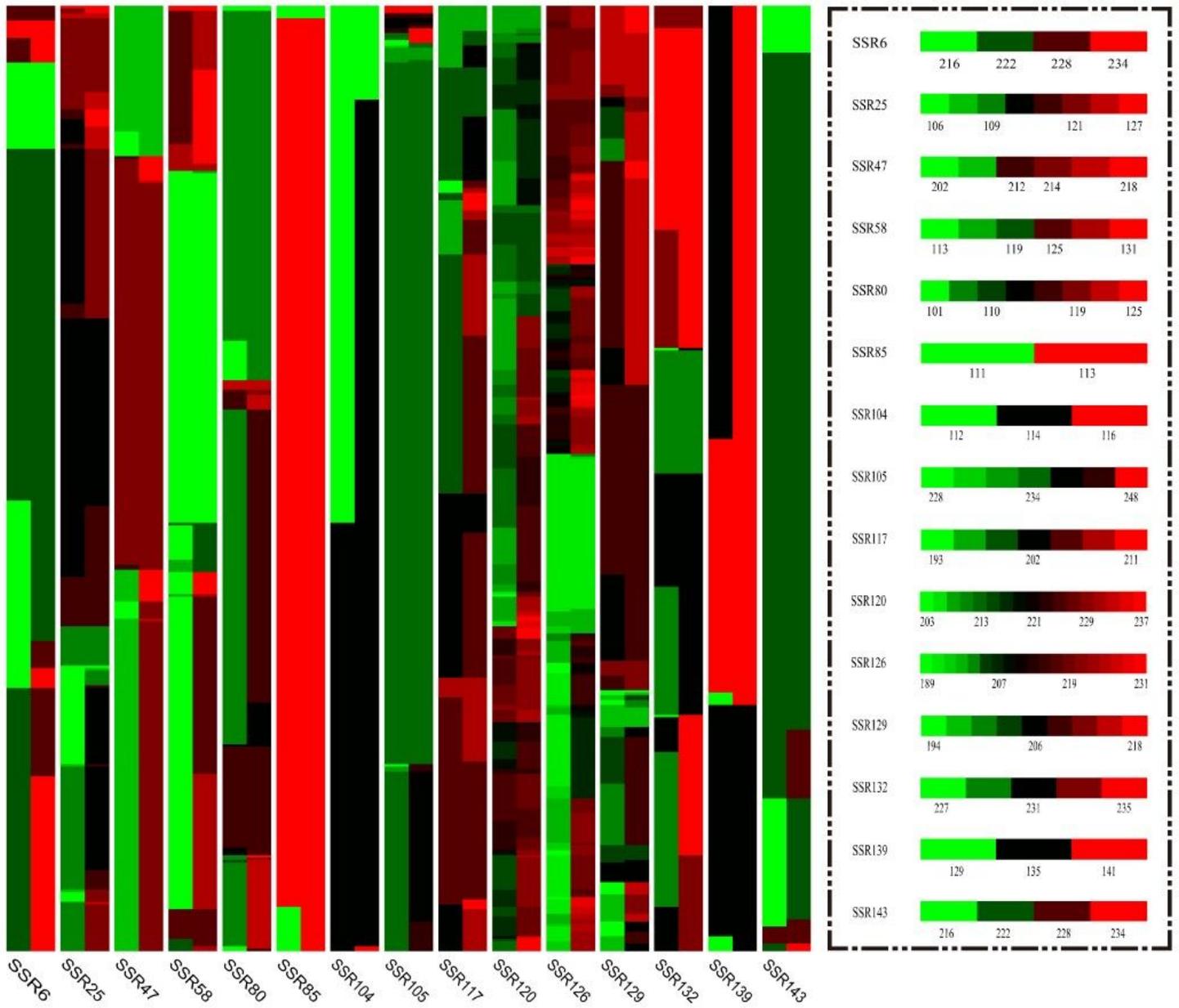
$H_o$  (pop.): mean of observed heterozygosity of population;  $H_e$  (pop.): mean of expected heterozygosity of population;  $F_{st}$ : inter-population genetic fraction coefficient;  $N_m$ : gene flow.

**Table 4. Gene flow ( $N_m$ , above diagonal) and genetic differentiation coefficient ( $F_{st}$ , below diagonal) between the six provenances**

	Mis	Iow	Was	Que	Lou	Ten
Mis	-	1.885	2.959	4.476	4.476	3.681
Iow	0.117	-	1.618	2.038	2.155	2.238
Was	0.078	0.134	-	2.666	2.056	1.785
Que	0.053	0.109	0.086	-	4.469	4.250
Lou	0.049	0.104	0.108	0.053	-	8.469
Ten	0.064	0.100	0.123	0.056	0.029	-

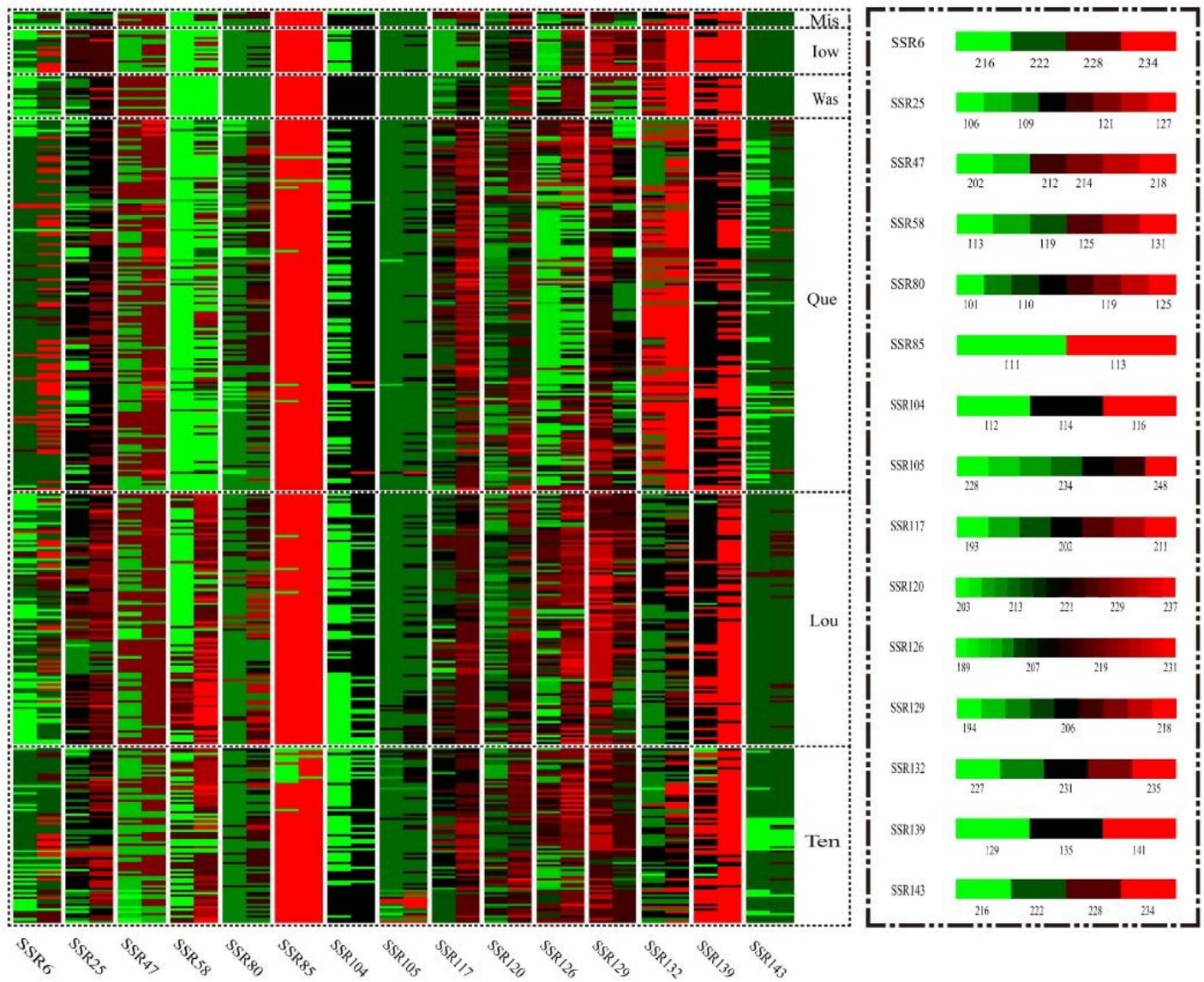
'Mis', 'Iow', 'Was', 'Que', 'Lou', 'Ten': different provenance populations respectively.  $F_{st}$ : inter-population genetic fraction coefficient;  $N_m$ : gene flow.

## Figures



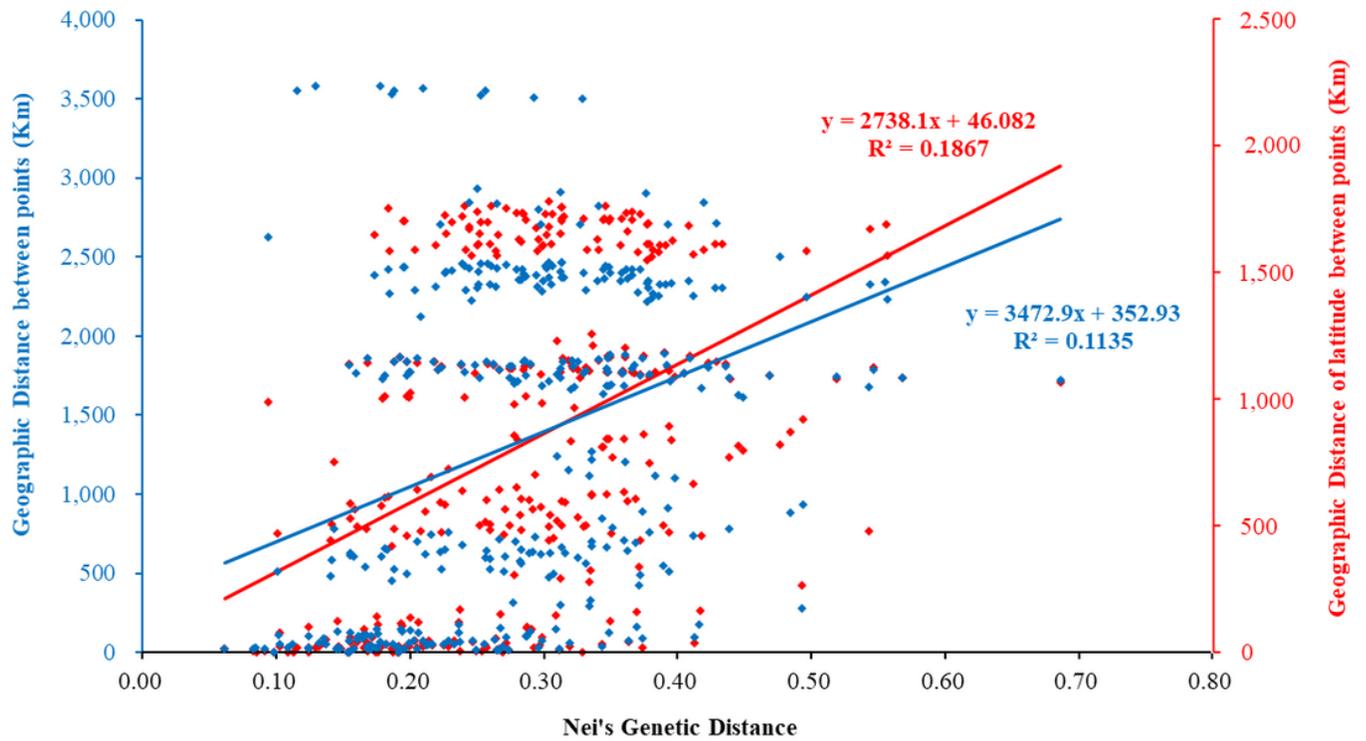
**Figure 1**

Heat map of the polymorphic SSR loci. The length of amplified fragments at each locus is represented by a colored band and each locus has a unique contrast diagram of fragment size and color. Different colors represent different fragment sizes, with greener colors indicating smaller fragments; redder colors indicating larger fragments.



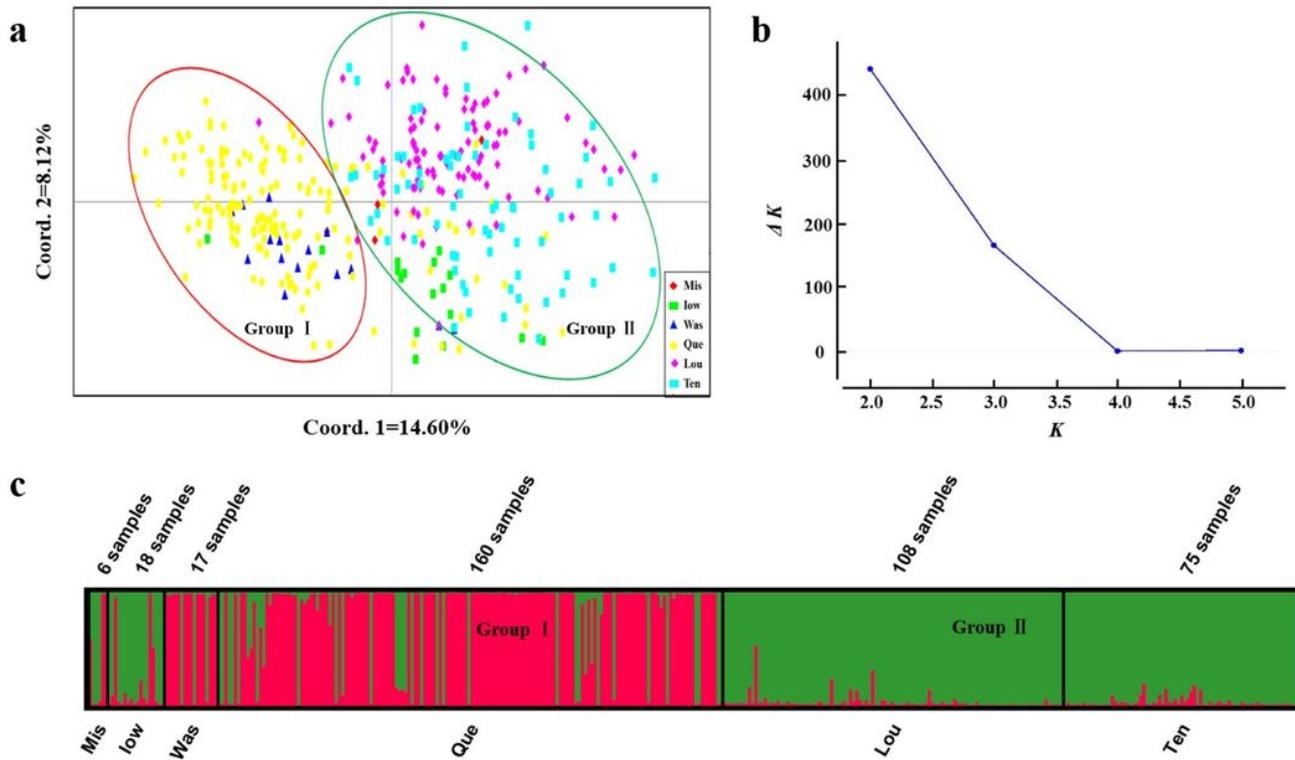
**Figure 2**

The heat map of population genetic diversity. The length of the amplified fragments at each locus is represented by a colored band and each locus has a unique contrast diagram of fragment size and color. Different colors represent different fragment sizes, a greener color indicates smaller fragments; a redder color indicates larger fragments. Individuals of the same population are surrounded by black dotted frames. 'Mis', 'low', 'Was', 'Que', 'Lou', 'Ten': different provenance populations respectively.



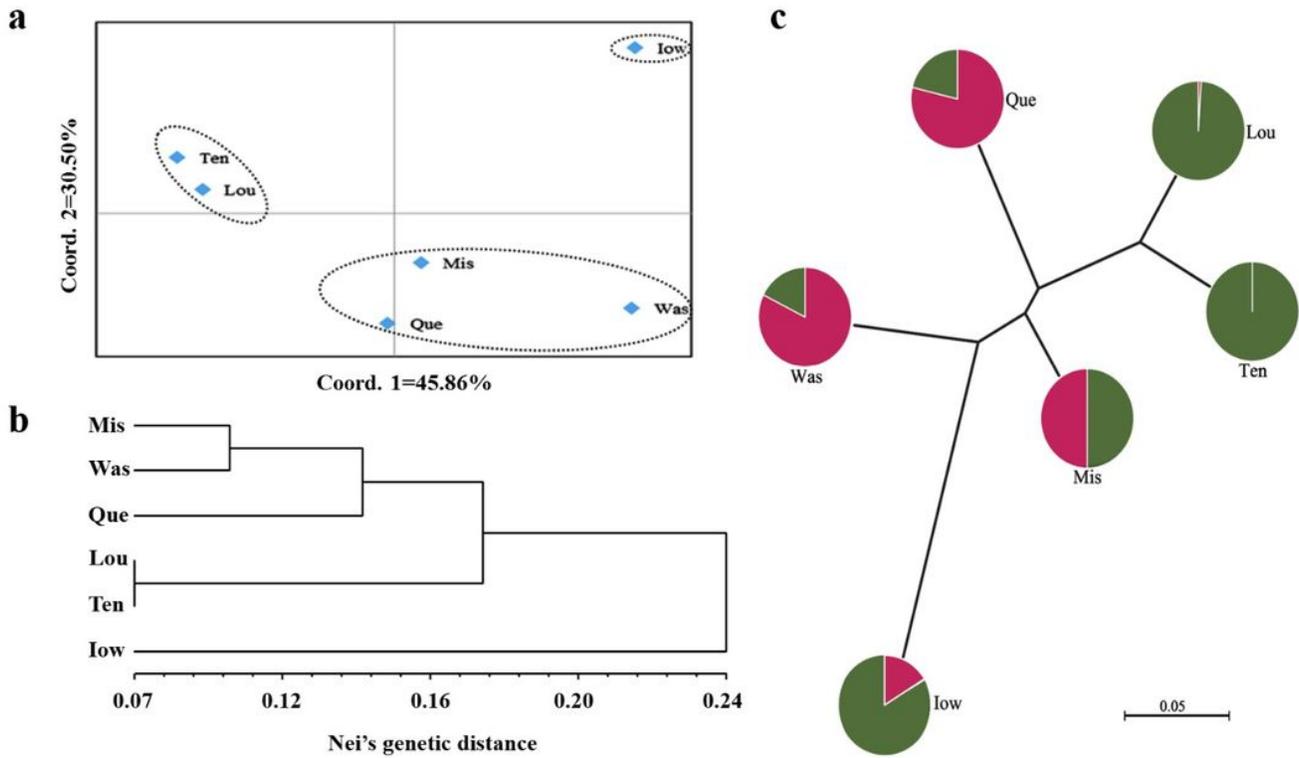
**Figure 3**

Correlation between Nei's genetic distance and geographical distance ( $P < 0.01$ ). The blue part represents the correlation between Nei's genetic distance and geographical distance between points, and the red part represents the correlation between Nei's genetic distance and latitude distance between points.



**Figure 4**

Analysis of the population structure of 384 *P. deltoides* Individuals. (a) Principal coordinates analysis (PCoA) of 384 individuals from six provenances. ◆: the provenance population in Missouri, USA ('Mis'); ■: the provenance population in Iowa, USA ('low'); ▲: the provenance population in Washington State, USA ('Was'); ●: the provenance population in Quebec, Canada ('Que'); ◆: the provenance population in Louisiana, USA ('Lou'); ■: the provenance population in Tennessee, USA ('Ten'). Red circle area: Group I; Blue circle area: Group II. (b) Relations between the number of K and  $\Delta K$ , based on the model developed by Evanno et al. [37]. (c) The population structure of *P. deltoides* determined using STRUCTURE 2.3.4 [38] software (K = 2). Red area: Group I; Blue area: Group II. 'Mis', 'low', 'Was', 'Que', 'Lou', 'Ten': different provenance populations respectively.



**Figure 5**

Population structure of six provenance populations of *P. deltoides*. (a) The principal coordinates analysis (PCoA) of six populations of *P. deltoides*. (b) UPGMA analysis of six populations of *P. deltoides* based on Nei's genetic distance. (c) The unrooted tree based on Nei's genetic distance for six *P. deltoides* populations. The pie chart reflects the distribution proportion of individuals of the provenance populations in the two groups.

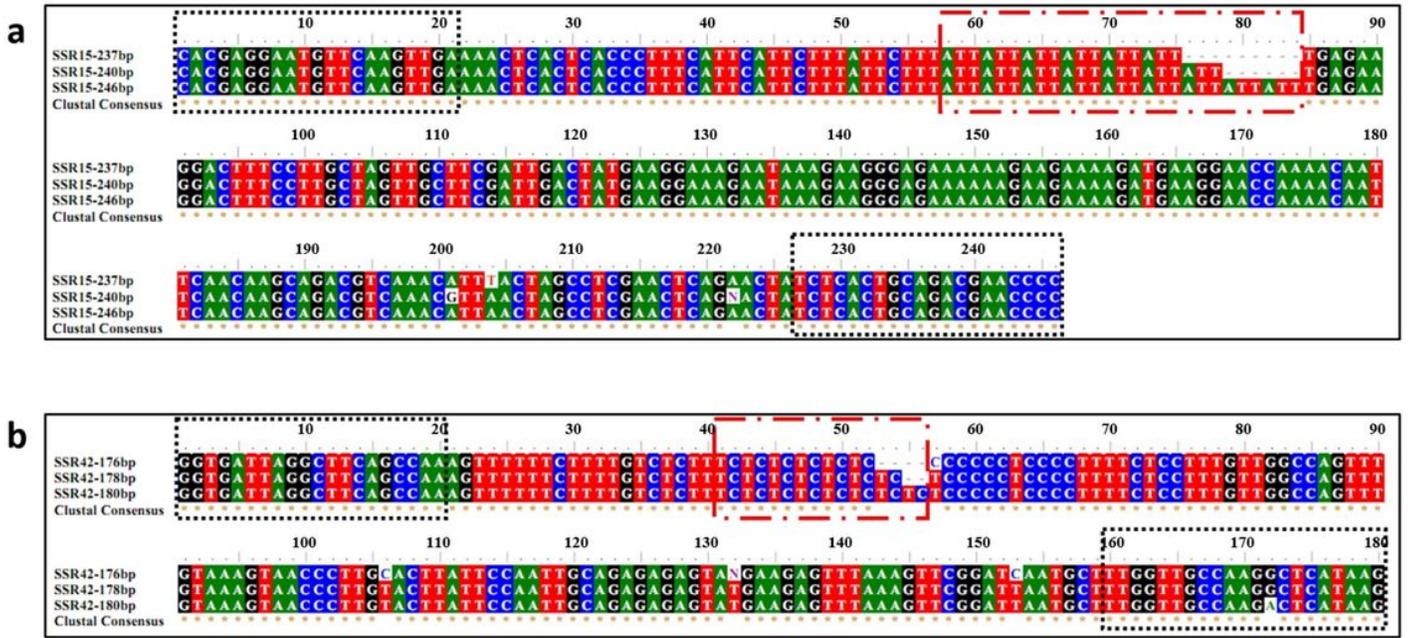


Figure 6

Amplification sequence alignment diagram of SSR15 (a) and SSR42 (b). The base sequence in the black dotted frame matches the corresponding primer sequence, the sequence in the red dotted frame is a simple repeat sequence

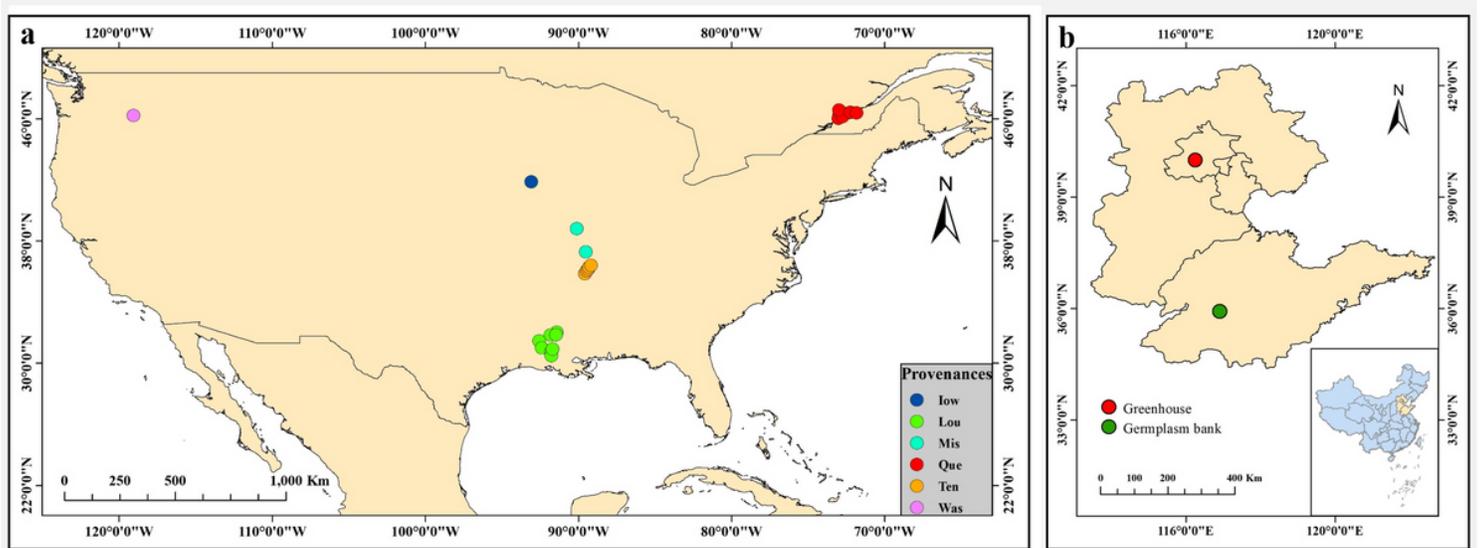


Figure 7

Geographical distribution of provenances (a) and germplasm bank (b) of *P. deltoides*

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS1.jpg](#)
- [TableS7.xlsx](#)
- [TableS4.docx](#)
- [TableS6.docx](#)
- [TableS1.docx](#)
- [TableS3.docx](#)
- [TableS2.docx](#)