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)
- 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: July 24th, 2019

DOI: https://doi.org/10.21203/rs.2.10562/v1

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 of P. deltoides. Ultimately, 108 alleles (Na) were detected; the expected heterozygosity (He) ranged from 0.070 to 0.905, and the average locus polymorphic information content (PIC) was 0.5354. The provenance ‘Was’ had the lowest 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 (Fst) and gene flow (Nm) 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, clones distributed in 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 smaller gene flow and higher degree of genetic differentiation compared with the other provenances, and was classified into one group. There was no 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; however, the genetic variation among provenances has no clear correlation with their geographical locations. Clones reflecting population genetic information can be selected to build a core resource bank, which could provide effective protection and promote the scientific utilization of the P. deltoids resource, laying a solid foundation for poplar breeding.

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 understanding 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. Moreover, many techniques and methods have been used to evaluate population genetics, for example molecular marker technology is widely used in this field. 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 [10-13]. In addition, similar studies have been carried out on other tree species, such as *P. nigra* L. [14-19], *P. simonii* [20], *P. tremuloides* [21, 22], *P. balsamifera* [23, 24], *P. cathayana* Rehd [25], *P. euphratica* [26, 27], *P. tomentosa* [28], and *P. szechuanicavar. tibetica* [29].

P. deltoides is widely distributed from the Mississippi River to southern Canada in North America, is one of the keystone forest tree species, and has 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 cultivar are mainly used to provide feedstocks for pulp, fiber, and bioenergy industries [30]. 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. Production practice showed that the black poplar has a dominant position in the poplar plantation: *P. deltoides* and its hybrid with *P. nigra* L. have become the main commercial poplar trees in China.

There have been few studies on genetic diversity and population structure of *P. deltoides*. Fahrenkrog *et al.* [31] studied the genetic information of *P. deltoides* distributed in the Mississippi River Basin. However, in China, there has been little systematic and comprehensive research on the genetic information of introduced *P. deltoides* resources. Therefore, we have collected the germplasm resources of *P. deltoides* introduced by our research team and established a resource bank containing more than 1000 individuals. This bank lays the foundation for our future research on genetic diversity and population structure.

Among the methods and techniques to evaluate genetic diversity and population structure, simple sequence repeat (SSR) molecular marker technology has been widely used and is an ideal method because of its simple operation, co-dominance, high resolution, polymorphism, and repeatability [32, 33].

In the present study, we screened 15 pairs of stable, polymorphic SSR primers to analyze the genetic diversity and population structure of 384 *P. deltoides*individuals that were randomly selected from six provenances in the gene pool. The aim was to understand the genetic information of the introduced *P. deltoides* resources, and to provide theoretical support for the construction of a core germplasm resource bank. These resources could then be protected, managed, and utilized more scientifically and rationally.

Results

Polymorphic SSR primers

In this study, 20 stable, polymorphic SSR primer pairs (8 pairs of primers for functional gene development and 12 pairs of universal primers) were selected from among 145 poplar SSR primer pairs for further analysis (Additional file 2: Table S1).

Microsatellite polymorphisms

The results of null allele detection revealed that some loci demonstrated an excess of homozygotes and that null alleles existed. The null allele frequency was identified and estimated using Cervus 3.0.7 [34], and those primers with a null allele frequency greater than 0.2 were discarded. Based on the intersection of the results of two detection methods (Additional file 2: Table S2), the detection data of five loci (SSR15, SSR42, SSR54, SSR65, and SSR76) were ignored in the subsequent analysis to ensure the accuracy of the results. However, five other loci (SSR47, SSR85, SSR105, SSR129, and SSR143) showed the presence of null alleles in only one provenance population, and we believed that they did not have a significant impact on the genetic diversity of the population.

A total of 108 alleles were detected using 15 SSR primer pairs in 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 number of effective alleles (N_e) per locus was 3.476, ranging from 1.076 at SSR85 to 10.397 at SSR126. Locus SSR126 provided more genetic information than the other loci and its Shannon information index was the largest (I = 2.533). The means of the observed heterozygosity(H_o) and expected heterozygosity (H_e) were 0.509 and 0.579, respectively, and except for locus SSR104, H_o was higher than H_e, indicating that the false-homozygous phenomenon existed in some loci. 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), possibly because of differences in the number and structure of the *P. deltoides* resources in each provenance, which indicted that there was genetic differentiation among the different provenances (Additional file 2: Table S3). The analysis results showed that the selected SSR primers had abundant polymorphisms and could effectively reflect the genetic diversity of the *P. deltoides* population.

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 more monotonous 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 2: Table S4), we found that the ‘Was’ population had the lowest genetic diversity (N_a = 2.600, N_e = 1.962, I = 0.617, H_o = 0.365, H_e = 0.358). The I values of the ‘Que’, ‘Lou’, and ‘Ten’ populations were greater than 1.0, and these three populations had private alleles; therefore, their genetic diversities were higher than the others. The HWE results showed that the ‘Que’ and ‘Ten’ populations deviated significantly from the equilibrium, suggesting that genetic differentiation might have occurred within the populations. The fixation index (F) reflects whether the actual frequency of a genotype deviates from the proportion of the genetic equilibrium theory. The F value of the ‘Mis’ population was 0.210 and its H_e (0.502) was higher than its H_o (0.389),

suggesting that there was a genetic deletion in this population, which may be related to the small number of samples contained in the population.

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 'Was' populations expressed as homozygote at 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 clones in the 'Que' and 'Ten' populations were deleted on the basis of the main amplified fragments; and some clones in the 'Lou' population were lengthened.

Population genetic differentiation and genetic variation

The *P. deltoides* 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, which was caused by a lack of heterozygosity from non-random mating among individuals. The inter-population genetic fraction coefficient (F_{st}) was used as an indicator to evaluate the level of genetic differentiation of populations: The larger the value, the greater the genetic differentiation among 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 [36]. The average of F_{is} and F_{it} values of *P. deltoides* 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 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. deltoides* was mainly caused by genetic differences among individuals.

Analysis using the Mantel test showed that there was no significant relationship between genetic distance and geographical distance ($R^2 = 0.0065$, $P = 0.333$, Fig. 3), indicating that geographical location was not the main factor leading to the genetic differences among the populations.

Genetic structure

The results of PCoA showed that the 384 clones of *P. deltoides* could be divided into two groups (Fig. 4a). Group I mainly included the clones from 'Que' and 'Was', and a few clones from the 'low' provenance. Group II mainly included the clones of the 'Ten', 'Lou', 'Mis,' and 'low' provenances. At the same time, the clones from the 'Que' provenance were widely distributed, which indicated that 'Que' has abundant genetic diversity. STRUCTURE cluster analysis showed that when $K = 2$, the ΔK value was the largest (Fig. 4b), which indicated that the 384 *P. deltoides* individuals were suitable to 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 2: Table S5).

To understand the genetic structure among the six populations of *P. deltoides*, 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 invalid alleles may reduce the population genetic diversity and increase genetic differentiation among populations, which will have a significant impact on the results of the study [22, 39, 40]. Therefore, when using molecular marker technology to carry out related research, the null alleles of the locus should be tested to ensure the accuracy and reliability of the results. In this study, two pieces of software with different calculation methods (Micro-Checker and Cervus) were used to detect the null alleles of the loci, and the commonly detected results were selected to improve the credibility of the results. Five pairs of primers with high frequency of null alleles were identified, and the data for these five loci were excluded in the subsequent analysis to ensure the reliability of the results.

To clearly understand the effect of loci containing null alleles on the population genetic analysis results, we compared the changes in of major genetic parameters before and after deletion of these loci. We found that the difference between H_o and H_e increased, the number of F_{st} increased, and the N_m decreased in the presence of the null allele loci (Table 3). These changes indicated that the null allele loci increase the false-homozygous phenomenon in populations, expand genetic differentiation, and weaken gene exchange among populations, thus concealing genetic diversity of the populations to some extent [41, 42].

SSR Primer Screening

The SSR primers used in this study were partly derived from functional gene primers developed by other researchers. The amplification motifs of the two primer pairs, SSR15 and SSR42, showed some differences from the original study.

The SSR15 locus was located in the promoter region of the *PsHsf16* gene of the *P. simonii* Hsf family. Its repeat unit was ATTT [43]. In this study, the repeat motif of the SSR15 primer 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 [44]. In this study, the amplifying motif of the SSR42 primer was TC (Fig. 6b). At the same time, by calculating the null allele frequencies of different loci, 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 phenomena indicate that the marker information and characteristics of the same primers may change when they are applied to different experimental materials [45]. Therefore, the stability and polymorphism of primers should be further verified when screening functional gene development primers. Of course, it is better to develop polymorphic loci independently, based on the experimental materials, 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 [46]. Studies on the genetic diversity of *P. deltoides* populations have been carried out. Fahrenkrog *et al.* performed the first population genomics study for 425 unrelated individuals of *P. deltoides* distributed in 13 states of the southeastern United States. By assessing population structure, population differentiation, genetic diversity, LD, and adaptation in *P. deltoides*, they found that the differentiation between subpopulations of the natural *P. deltoides* population was weak (F_{st} = 0.022-0.106); however, the genetic diversity was high [31]. Furthermore, genome-wide association studies (GWASs) were also used to dissect the genetic regulation of eight growth and wood composition traits in *P. deltoides*, 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 [47]. In the present study, the genetic diversity and population structure of 384 individuals of *P. deltoides* 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 population had abundant genetic diversity and the average of Shannon's information index (*I*) was 0.910. Moreover, there was moderate genetic differentiation (F_{st}= 0.062-0.205, mean = 0.129), which was higher than the results of Fahrenkrog *et al.* [31]. This phenomenon may have been caused by the wide distribution of the clones selected in this study, which can more effectively reflect the genetic

information of *P. deltoides* populations. In addition, our AMOVA results showed that 84.88% of the genetic variation originated from individuals, and the difference between populations was small. Based on the results of the present study, we can build a core germplasm resource bank to rationally manage, preserve, and utilize the introduced *P. deltoides* resources.

Population structure

The *P. deltoides* population has high genetic diversity and moderate genetic differentiation. To further understand its population structure, association studies in natural populations are required [48]. In this study, we analyzed the population structure of *P. deltoides* based on different levels of analysis. At the individual level, PCoA and STRUCTURE cluster analysis were performed on 384 clones, which divided them into two groups. Most of the clones distributed in the Mississippi River Basin were grouped together (Group II, Fig. 4a, c), which was consistent with the results of Fahrenkrog *et al*'s. [31, 47] analysis. The genetic differentiation between different states in the Mississippi River Basin is weak; however, it is larger than other basins, indicating that geographical isolation restricts gene exchange among populations, a conclusion reached by certain other researchers [11, 49]. By contrast individuals distributed in the Saint Lawrence River Basin and the Columbia River Basin, with similar climatic conditions and latitudes, were grouped into the other group (Group I, Fig. 4a, c). Clones distributed between the two locations may have the same type of genetic variation because of the similarities between the external environments, or there may be introducing behavior between the two provenances. Taking the population as the research object, PCoA and UPGMA cluster analysis were carried out, and the unrooted trees among the provenance populations were plotted. The results showed that 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). Meanwhile, there was no correlation between genetic distance and geographical distance among populations. To determine the reasons for the differences between the results of individual and population analysis, we analyzed N_m and F_{st} among the six provenance populations (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 [50, 51]. Eighteen individuals from 'low' were introduced, and their H_o (0.456) was significantly higher than the H_e (0.390) (Additional file 2: Table S4). Moreover, PCoA of 384 individuals of *P. deltoides* showed that 18 clones from 'low' distributed widely in the dimension of the first principal component, and most clones were on the edge of Group II (Fig. 4a). Therefore, we have two different hypotheses. One is that there may be more clones of *P. deltoides* in 'low', but in this study, only 18 introduced clones were studied. The genetic information reflected by these clones differed greatly from that of other provenances. However, these clones were reasonably classified into one group at the individual level. The other hypothesis is that individuals from 'low' may be subjected to some external selection pressure, which has resulted in genetic variation of this *P. deltoides* population in a certain direction. This selection pressure may exist in other provenances, but the selection pressure is relatively weak. Therefore, in the analysis of population structure, we should not neglect the influence of external environmental conditions on the structure.

Methods

Plant materials and DNA extraction

The cuttings of *P. deltoides* were collected from the germplasm resource bank that established through international exchange and cooperation of germplasm resources, 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. A total of 384 clones were selected in this research. Among them, 6, 18, 75, and 108 individuals were from Missouri ('Mis'), Iowa ('Iow'), Tennessee ('Ten'), and Louisiana ('Lou'), respectively. The four provenances were located in Mississippi River basin. Seventeen individuals were from Washington ('Was'), which is located in the Columbia River basin. In addition, 160 clones were from the Saint Lawrence River basin (Quebec, Canada, 'Que') (Additional file 1: Fig. S1, Additional file 2: Table S6).

Plant 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 [52]. The quality and integrity of the extracted genomic DNA were assessed using 1% agarose gel electrophoresis. The concentration was then determined using a NanoDrop-2000 ultramicro-spectrophotometer (Thermo Fisher Scientific Waltham, MA, USA). Finally, the DNA was diluted to 50 ng·μL^{-1} and stored at -20 C for polymerase chain reaction (PCR) amplification.

Primer selection and PCR amplification

In this study, 20 representative clonal DNAs of *P. deltoides* were selected randomly for SSR primer screening, which included 75 pairs of primers for functional gene development of poplar [43, 44, 53-56] and 70 pairs of universal primers [57-60]. Using PCR amplification, 2% agarose gel electrophoresis, and 8% non-denaturing polyacrylamide gel electrophoresis, the SSR primers were screened for their stability and polymorphism content.

The PCR amplification reaction system for all SSR markers comprised a 25-μL mixed system containing 2.5 μL of 10 × buffer (Mg^{2+} plus), 1.8 μL of dNTP mixture, 1 μL of forward primer (10 μmol·L^{-1}), 1 μL of reverse primer (10 μmol·L^{-1}), 0.25 μL of *Taq* polymerase (5 U·μL^{-1}), 1 μL of DNA template (50 ng·μL^{-1}), and 17.45 μL of ddH_2O. 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 PCR products of the polymorphic SSR primers were separated by capillary electrophoresis using an ABI 3730xl DNA analyzer (Applied Biosystems, Foster City, CA, USA) and the peak size map of 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. To exclude the effect of null alleles on the population genetic diversity, the null allele frequencies of each locus were detected using Cervus 3.0.7 [34] and Microchecker 2.2.3 [35] software, and the Cervus 3.0.7 software was also used to calculate *PIC* of each locus.

The GeneAlEx 6.503 [61, 62] software was used to convert various file formats for different analysis and to calculate genetic diversity parameters, including the number of observed alleles (N_a), the number of effective alleles (N_e), Shannon's information index (I), observed heterozygosity (H_o), expected heterozygosity (H_e), the number of private alleles, population fixed index (F), gene flow (N_m), and *F*-Statistics (F_{is}, F_{it}, F_{st}). HWE tests across all populations were performed using Genepop 4.7.0 [63, 64]. To more intuitively reflect the genetic diversity of loci and populations, heat maps of every locus were drawn using OmicShare tools (www.omicshare.com/tools). With the help of the Arlequin 3.5.2.2 [65] and GeneAlEx 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 GeneAlEx 6.503 software for PCoA. A cluster analysis between populations, based on UPGMA, was also developed using the NTSYS-pc 2.10e [66] software. An unrooted tree was constructed based on pairwise standard genetic distances [67], using the least squares algorithm with 10,000 bootstrap replicates, and these processes were generated and analyzed using PHYLIP 3.6 [68] software.

The population genetic structure was analyzed using STRUCTURE 2.3.4 [38] 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 1000,000 MCMC iterations. The ΔK 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.* [37].

To determine whether further geographical distance can cause genetic differences, a Mantel test was performed using the GeneAlEx 6.503 software to analyze the correlation between Nei's genetic distance and geographical distance. The geographical distance between different provenances was calculated according to the latitude and longitude using Vincenty's formula (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; 'Iow': 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].

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 XS conceived and designed the experiments. CC 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, Prinz K, Kleemann F, Köhler L, *et al.* Genetic diversity in aspen and its relation to arthropod abundance. Front Plant Sci. 2015;5:806.

[2] Haddad NM, Crutsinger GM, Gross K, Haarstad J, Tilman D. Plant diversity and the stability of foodwebs. Ecology Letters. 2011;14(1):42-6.

[3] Bailey JK, Schweitzer JA, Ubeda F, Koricheva J, LeRoy CJ, Madritch MD, *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, Mcinnes KH. 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, Ma T, Uematsu C, Katayama H. 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, Eckert AJ, Legendre P, Neale DB. 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, Bohlmann J, Grigoriev I, Hellsten U, *et al.*The genome of black cottonwood, *Populus trichocarpa* (Torr & Gray). Science. 2006;313(5793):1596-604.

[10] Slavov GT, DiFazio SP, Martin J, Schackwitz W, Muchero W, Rodgers-Melnick E, *et al.* Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa*. The New phytologist.2012;196(3):713-25.

[11] Zhou L, Bawa R, Holliday JA. Exome resequencing reveals signatures of demographic and adaptive processes across the genome and range of black cottonwood (*Populus trichocarpa*). Molecular Ecology. 2014; 23(10):2486-99.

[12] Evans LM, Slavov GT, Rodgers-Melnick E, Martin J, Ranjan P, Muchero W, *et al.* Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. Nature Genetics. 2014;46(10):1089-96.

[13] Holliday JA, Zhou L, Bawa R, Zhang M, Oubida RW. 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.

[14] Jiang DC, Wu GL, Mao KS, Feng JJ. Structure of genetic diversity in marginal populations of black poplar (*Populus nigra* L). Biochemical Systematics and Ecology. 2015;61:297-302.

[15] Guet J, Fabbrini F, Fichot R, Sabatti M, Bastien C, Brignolas F. 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.

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

[17] Fossati T, Grassi F, Sala F, Castiglione S. Molecular analysis of natural populations of *Populus nigra* L intermingled with cultivated hybrids. Molecular Ecology. 2010;12(8):2033-43.

[18] Schoot JVD, Pospíšková M, Vosman B, Smulders, MJM. Development and characterization of microsatellite markers in black poplar (*Populus nigra* L). Theoretical and Applied Genetics. 2000;101(1-2):317-22.

[19] Zhang YX, Zhang XR, Che Z, Wang LH, Wei WL, Li DH. 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.

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

[21] Callahan CM, Rowe CA, Ryel RJ, Shaw JD, Madritch MD, Mock KE. Continental-scale assessment of genetic diversity and population structure in quaking aspen (*Populus tremuloides*). Journal of Biogeography. 2013;40(9):1780-91.

[22] Woody JD, Rickman TH, Jones BE, Hipkins VD. Allozyme and microsatellite data reveal small clone size and high genetic diversity in aspen in the southern Cascade Mountains. Forest Ecology and Management. 2009;258(5):687-96.

[23] Keller SR, Olson MS, Silim S, Schroeder W, Tiffin P. Genomic diversity, population structure, and migration following rapid range expansion in the Balsam Poplar, *Populus balsamifera*. Molecular Ecology. 2010;19(6):1212-26.

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

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

[26] Saito Y, Shiraishi S, Tanimoto T, Yin L, Watanabe S, Ide Y. Genetic diversity of *Populus euphratica* populations in northwestern China determined by RAPD DNA analysis. New Forests. 2002;23(2):97-103.

[27] Wang J, Li Z, Guo Q, Ren GP, Wu YX. 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.

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

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

[30] 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.

[31] Fahrenkrog AM, Neves LG, Resende Jr MF, Dervinis C, Davenport R, Barbazuk WB, *et al.* Population genomics of the eastern cottonwood (*Populus deltoides*). Ecology and evolution. 2017;7(22):9426-40.

[32] Blair M, González L, Kimani P, Butare L. Genetic diversity, inter-gene pool introgression and nutritional quality of common beans (*Phaseolus vulgaris* L) from Central Africa. Theor Appl Genet. 2010;121:237-48.

[33] Ebrahimi A, Zarei A, McKenna JR, Bujdoso G, Woeste KE. Genetic diversity of Persian walnut (*Juglans regia*) in the cold-temperate zone of the United States and Europe. Sci Hortic. 2017;220:36-41.

[34] 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.

[35] Oosterhout CV, Hutchinson WFD, Wills D, Shipley P. Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes, 2004;4(3):535-8.

[36] Slatkin M. Gene flow and the geographic structure of natural populations. Science. 1987;236(4803):787-92.

[37] 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.

[38] Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945-59.

[39] Hoffman JI, Amos W. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. Molecular Ecology. 2010;14(2):599-612.

[40] Bonin A, Bellemain E, Eidesen PB, Pompanon F, Taberlet P. How to track and assess genotyping errors in population genetic studies. Molecular Ecology. 2004;13(11):3261-73.

[41] Bruford MW, Ciofi C, Funk SM. Characteristics of microsatellites In Molecular Tools for Screening Biodiversity. Dordrecht: Springer; 1998. p. 202-5.

[42] Guo Q, Li XY, Yang SH, Yang ZH, Sun YH, Zhang JT, *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.

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

[44] 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.

[45] Ellegren H, Galtier N. Determinants of genetic diversity Nature Reviews Genetics. 2016;17:422-33.

[46] Porth I, Klápště J, McKown AD, Mantia JL, Guy RD, Ingvarsson PK, Hamelin R, Mansfield SD, Ehlting J, Douglas CJ, El-Kassaby YA. Evolutionary quantitative genomics of *Populus trichocarpa*. PloS one. 2015;10(11):e0142864.

[47] Fahrenkrog AM, Neves LG, Resende MF Jr, Vazquez AI, de Los Campos G, Dervinis C, *et al.* Genome-wide association study reveals putative regulators of bioenergy traits in *Populus deltoides*. New Phytologist. 2016;213(2):799.

[48] King RA, Harris SL, Karp A, Barker JHA. Characterisation and inheritance of nuclear microsatellite loci for use in population studies of the allotetraploid *Salix alba-Salix fragilis* complex. Tree Genetics & Genomes. 2010;6(2):247-58.

[49] Geraldes A, Farzaneh N, Grassa CJ, Mckown AD, Guy RD, Mansfield SD, *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.

[50] Bothwell HM, Cushman SA, Woolbright SA, Hersch-Green EI, Evans LM, Whitham TG, *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.

[51] Vranckx GU, Jacquemyn H, Muys B, Honnay O. Meta-analysis of susceptibility of woody plants to loss of genetic diversity through habitat fragmentation. Conservation Biology. 2012;26:228-37.

[52] Doyle. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987;19:11-5.

[53] Du QZ, Wang BW, Wang BL, Zhang M, Li BL, Zhang ZY, *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.

[54] Wang BL, Wang BW, Chen QQ, Li BL, Zhang DQ. Identification of SSR loci from transcription factor genes expressed under abiotic stresses in poplar. Scientia Silvae Sinicae. 2011;47(8):67-74.

[55] Du QZ, Pan W, Xu BH, Li BL, Zhang DQ. 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.

[56] Wei ZZ, Du QZ, Guo Q, Zhang JF, Li BL, Zhang DQ. *DREB* gene and its application in analyzing population structure in *Populus simonii*. Bulletin of Botany. 2010;45(5):556-65.

[57] Huang LJ, Su XH, Zhang XH, Huang QJ. SSR molecular markers related to wood density and fibre traits in poplar. Journal of Genetics and Genomics. 2004;31(3):299-304.

[58] Zhang XH, Su XH, Huang QJ, Zhang BY. Comporison of genetic variations of black poplar (*Populus nigra* L) gene resource by microsatellite markers. Forest Research. 2006;19(4):477-83.

[59] Wei ZZ, Zhang GY, Du QZ, Zhang JF, Li BL, Zhang DQ. Association mapping for morphological and physiological traits in *Populus simonii*. BMC Genetics. 2014;15(S1):S3.

[60] Politov DV, Belokon MM, Belokon YS, Polyakova TA, Shatokhina AV, Mudrik EA, *et al.* Application of microsatellite loci for molecular identification of elite genotypes, analysis of clonality, and genetic diversity in Aspen *Populus tremula* L (Salicaceae). Int J Plant Genomics. 2015;3:1-11.

[61] Peakall R, Smouse PE. GENALEX 6: genetic analysis in Excel Population genetic software for teaching and research. Molecular Ecology Notes. 2006;6:288-95.

[62] Peakall R, Smouse PE. GenAlEx 6.5: genetic analysis in Excel population genetic software for teaching and research an update. Bioinformatics. 2012;28:2537-9.

[63] Raymond M, Rousset F. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J. Heredity, 1995;86:248-9.

[64] Rousset F. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Molecular Ecology Resources. 2008;8:103-6.

[65] 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.

[66] Rohlf F J. NTSYS pc2.1: Numerical taxonomy and multivariate analysis system version 2.1. New York: Applied Biostatistics Inc. 2000.

[67] Nei M. Genetic Distance between Populations. American Naturalist. 1972;106(949):283-92.

[68] Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences. Seattle: University of Washington. 2005.

Figures

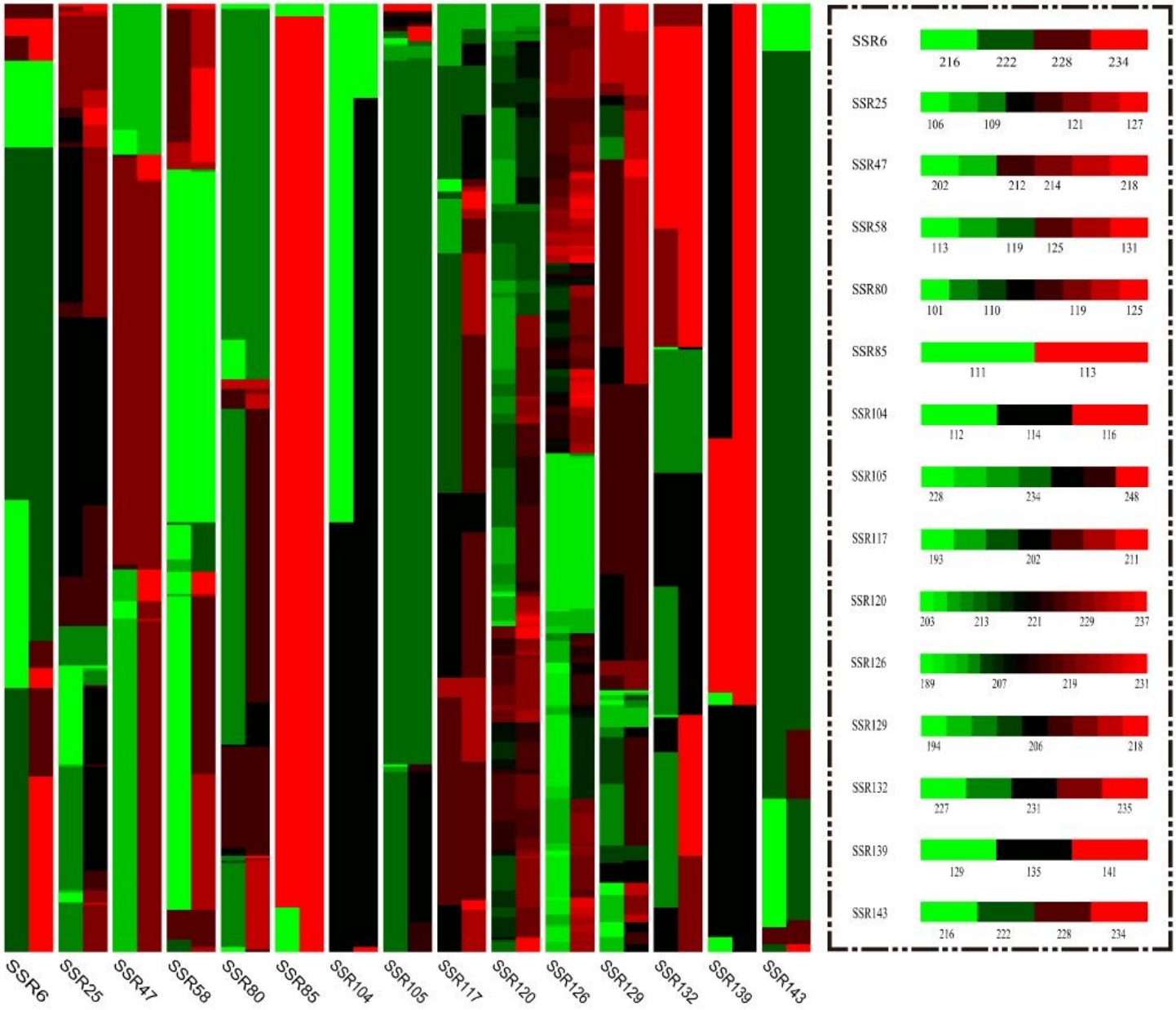

Figure 1

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.

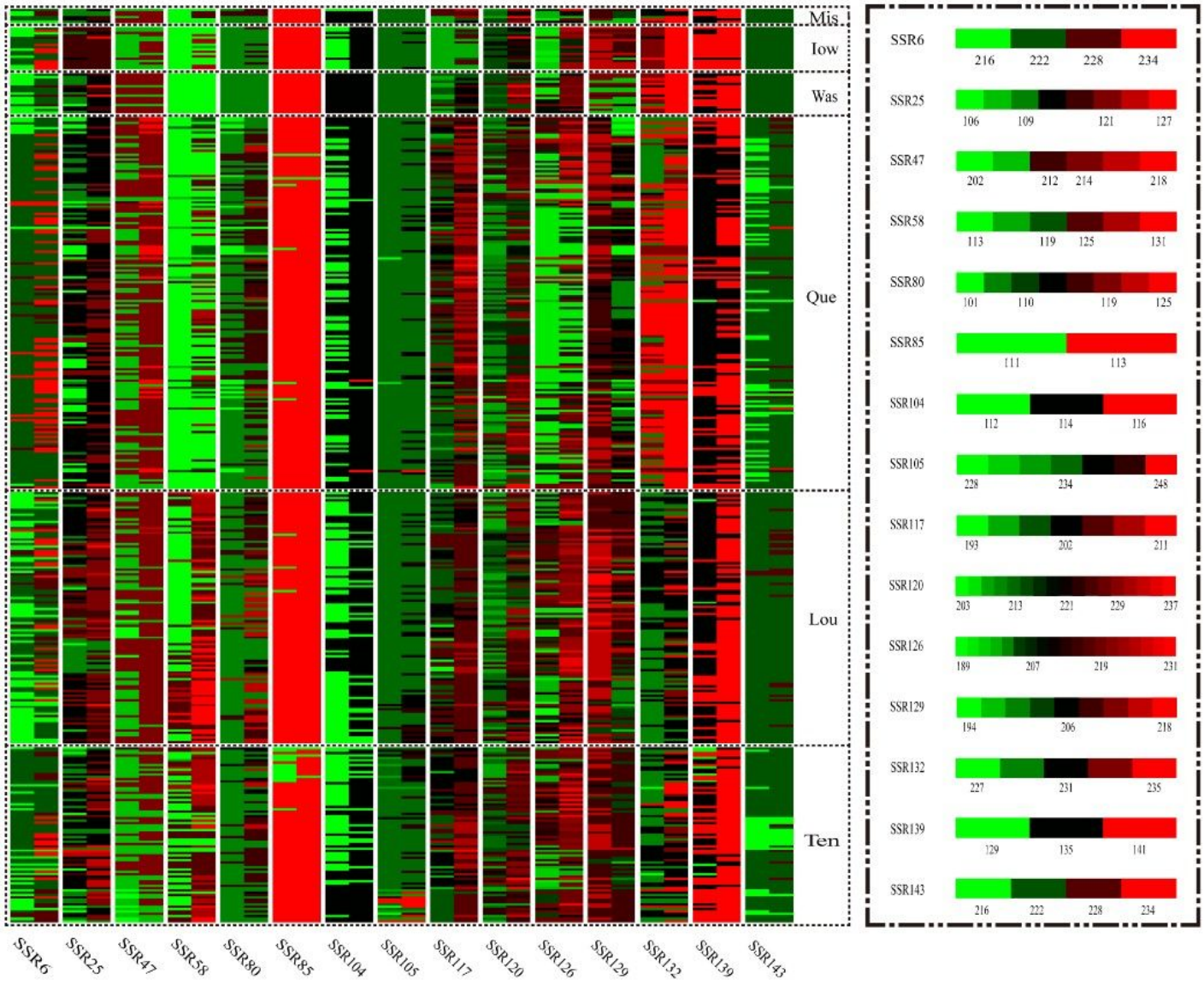

Figure 2

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.

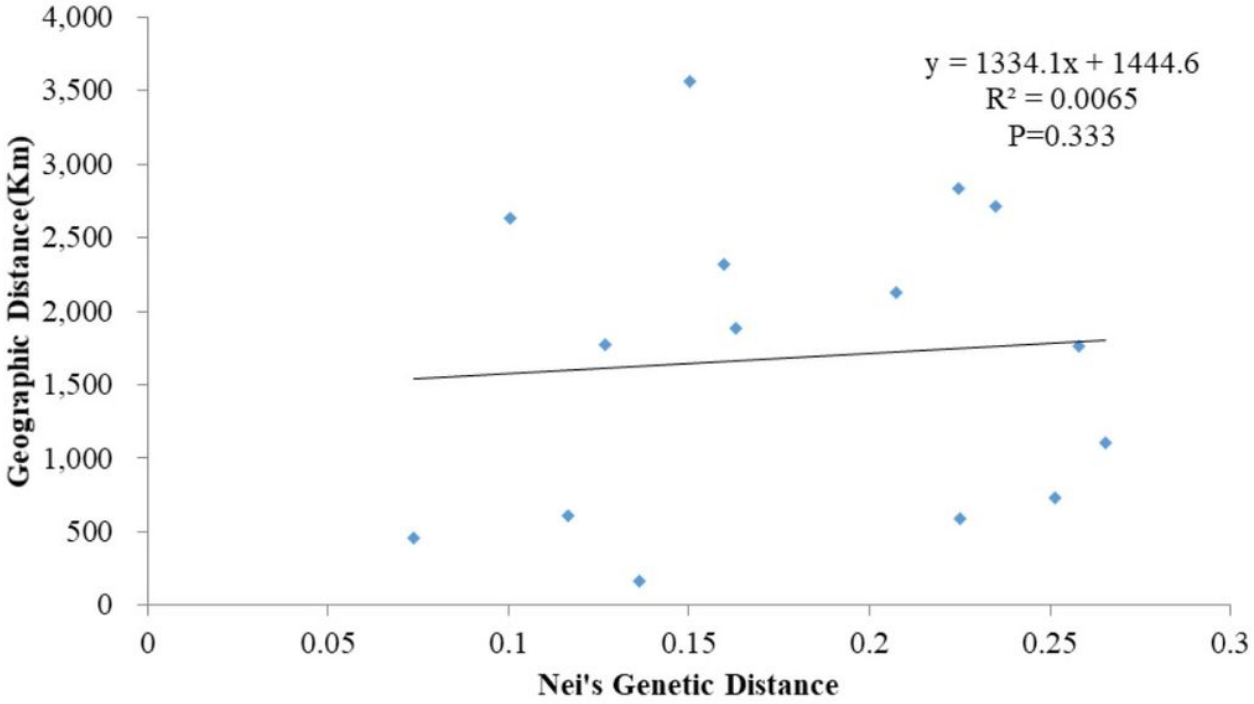

Figure 3

Correlation between Nei's genetic distance and geographical distance (P > 0.05)

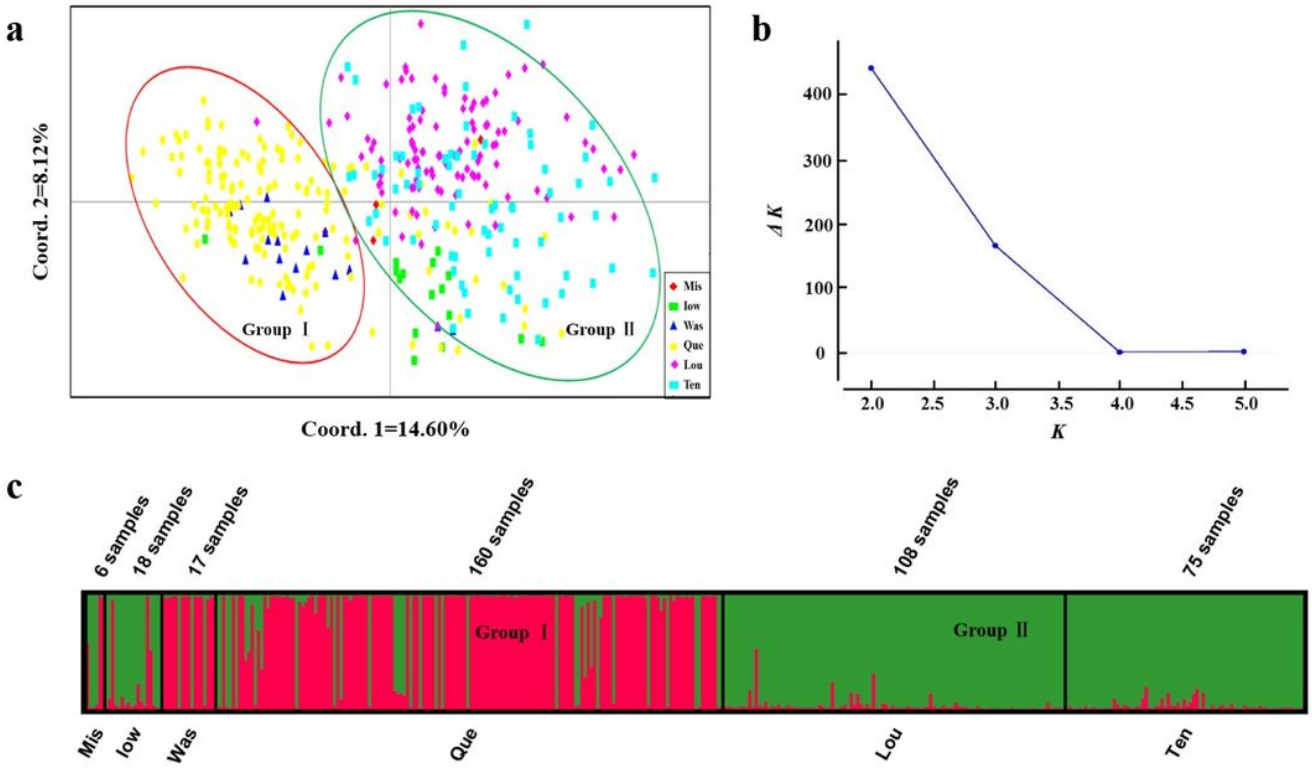

Figure 4

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 ('Iow'); ▲: 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 ΔK, 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', 'Iow', 'Was', 'Que', 'Lou', 'Ten': different provenance populations respectively.

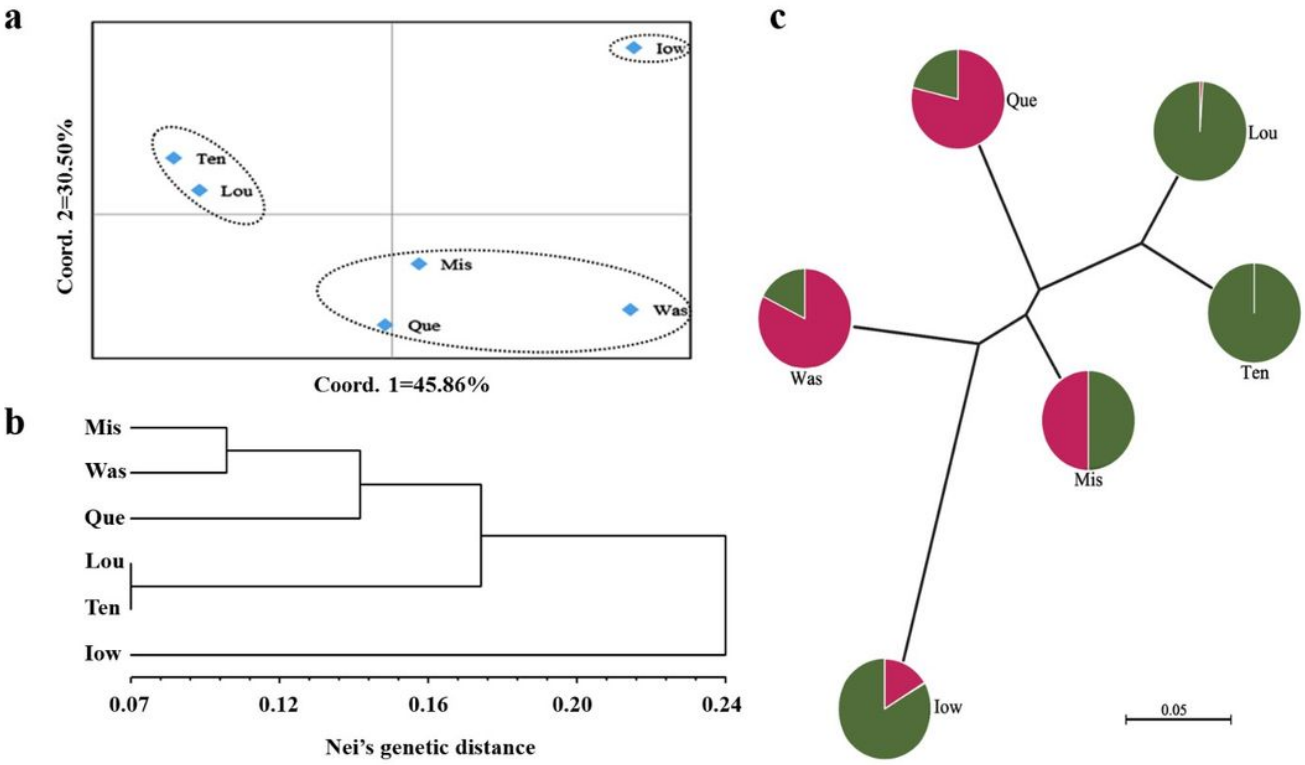

Figure 5

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.

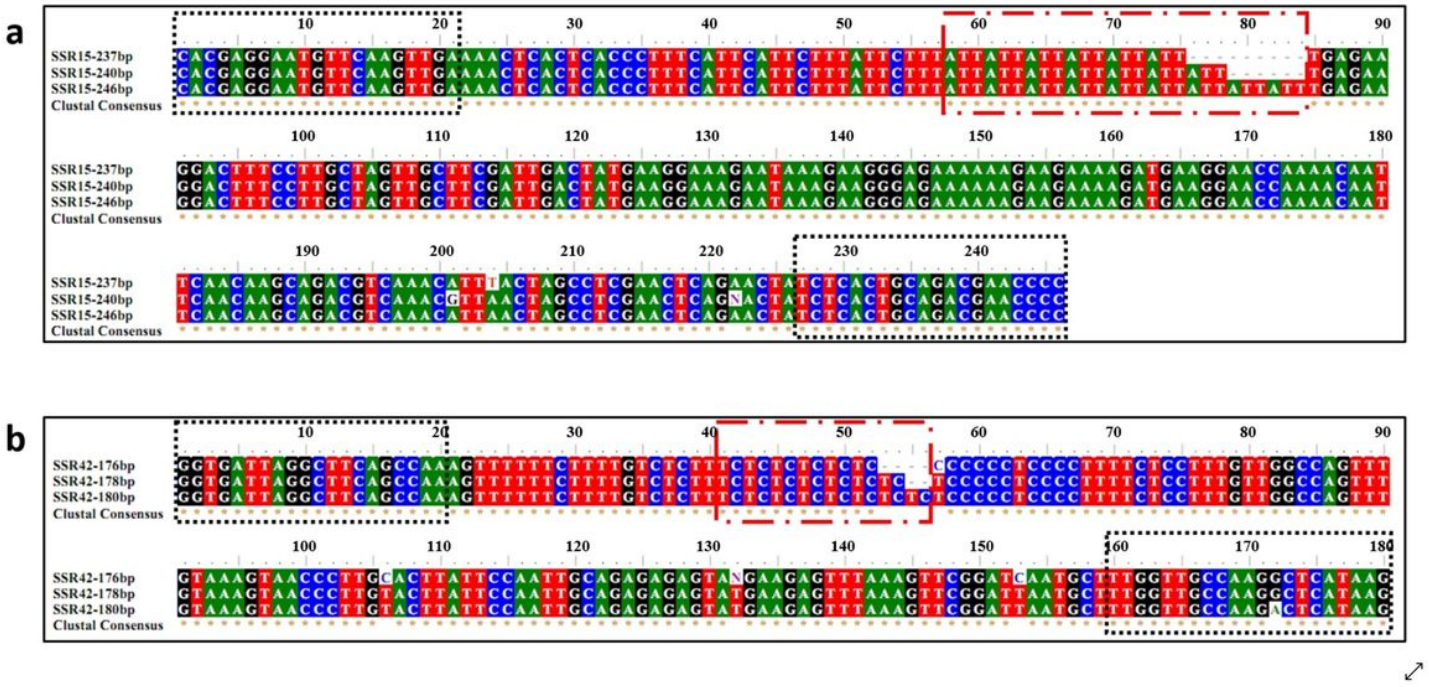

Figure 6

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

Supplementary Files

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

- supplement1.docx
- supplement1.docx
- supplement3.docx
- supplement3.docx
- supplement4.jpg
- supplement6.docx
- supplement7.docx