Electronic Journal of Biotechnology 26 (2017) 33–39

Contents lists available at ScienceDirect

Electronic Journal of Biotechnology

Research Article

Analysis of the genetic relationships and diversity among 11 populations of *Xanthoceras sorbifolia* using phenotypic and microsatellite marker data

Zhan Shen [a,b], Kaiquan Zhang [a,b], Luyi Ma [a,b,*], Jie Duan [a], Yan Ao [b]

[a] National Energy R&D Center for Non-food Biomass, Beijing Forestry University, Beijing 100083, PR China
[b] Key Laboratory for Silviculture and Conservation, Ministry of Education, Beijing Forestry University, Beijing 100083, PR China

ARTICLE INFO

Article history:
Received 22 August 2016
Accepted 29 December 2016
Available online 31 December 2016

Keywords:
Conservation of plant genetic resource
Genetic diversity
Germplasm characterization
Germplasm exploitation
Heterozygosity
Kernel oil content
Management of plant genetic resource
Microsatellite markers
Molecular markers
Polymorphism information content
Xanthoceras sorbifolia

ABSTRACT

Background: Assessments of genetic diversity are essential for germplasm characterization and exploitation. Molecular markers are valuable tools for exploring genetic variation and identifying germplasm. They play key roles in a *Xanthoceras sorbifolia* breeding program.
Results: We analyzed the genetic diversity of populations of this species using 23 simple sequence repeat (SSR) loci and data on kernel oil content. The 11 populations included in the study were distributed across a large geographic range in China. The kernel oil content differed significantly among populations. The SSR marker analysis detected high genetic diversity among the populations. All SSRs were polymorphic, and we identified 80 alleles across the populations. The number of alleles at each locus ranged from two to six, averaging 3.48 per primer pair. The polymorphism information content values ranged from 0.35 to 0.70, averaging 0.51. Expected heterozygosity, observed heterozygosity, and Shannon's information index calculations detected large genetic variations among populations of different provenance. The high average number of alleles per locus and the allelic diversity observed in the set of genotypes analyzed indicated that the genetic base of this species was relatively wide. The statistically significant positive correlation between genetic and geographic distances suggested adaptations to local conditions.
Conclusions: Microsatellite markers can be used to efficiently distinguish *X. sorbifolia* populations and assess their genetic diversity. The information we have provided will contribute to the conservation and management of this important plant genetic resource.

1. Introduction

Xanthoceras sorbifolia Bunge is a member of the family Sapindaceae. It is a small tree that produces edible fruits and seeds with high oil content. The plants are long living (up to 1000 years) and can tolerate drought, low temperature, alkaline soils, and low fertility. Extensive cross-pollination in the species produces high levels of heterozygosity and genetic polymorphism [1]. The oil from this plant has considerable potential for the production of biodiesel [2,3].

Improving the cultivation efficiency of this long-living species for oil production will require the development of accurate methods for identifying wild populations. In the past, *X. sorbifolia* was categorized according to morphological traits such as the features of the leaves, flowers, and fruits [4,5,6]. However, these characteristics are strongly influenced by environmental factors [7], and characterizations based upon them have created confusion in the research community. Molecular techniques based on DNA markers have proven much more reliable for genetic characterization. The techniques include random amplification of polymorphic DNA (RAPD) [8], amplified fragment length polymorphism analyses [9], and inter-simple sequence repeat (ISSR) analyses [10]. Microsatellites (simple sequence repeats; SSRs) have become one of the most popular molecular markers in many diverse fields [11,12,13]. SSRs have proven extremely valuable in the linkage mapping of many plant species [14,15,16]. Compared with other markers, SSRs have high polymorphism and information content, are reproducible, exhibit codominant inheritance, and have a wide distribution along the lengths of the chromosomes. This combination of advantages has made SSRs very popular in the analyses of many species, including *Jatropha curcas* L. [17], *Olea europaea* L. [18], *Liriodendron chinense* (Hemsl.) Sarg [19,20]., and *Vitis vinifera* L. [21,22].

X. sorbifolia is widely distributed in northern China, where it is an ideal species for biodiesel production. However, environmental conditions and the kernel oil content vary widely across the distribution range.

* Corresponding author at: National Energy R&D Center for Non-food Biomass, Beijing Forestry University, Beijing 100083, PR China.
E-mail address: maluyi@bjfu.edu.cn (L. Ma).
Peer review under responsibility of Pontificia Universidad Católica de Valparaíso.

http://dx.doi.org/10.1016/j.ejbt.2016.12.008

Unraveling the genetic diversity of this species requires the collection and management of its germplasm and the identification of diverse genotypes that may contribute to improved yield, i.e., better qualities and quantities of the oil product. Few SSR markers have been developed for *X. sorbifolia* [23], and they have not been used in the genetic studies of the species. In this study, we used SSR markers to examine the genetic variability of the species in wild populations in different regions of China (Fig. 1) for the accurate assessment of genetic diversity and intra-specific relationships. We also analyzed the relationship among genetic distance, geographic provenance, and kernel oil contents. On this basis, we provide guidelines and a scientific framework for better the exploration of the germplasm resources of the species.

2. Material and methods

2.1. Plant material

We collected 142 individuals from 11 wild populations located in different regions of China (Fig. 1). The geographical and environmental parameters at the 11 sites are listed in Table 1. The regions sampled were located between 34.25° and 44.87°E and between 87.97° and 129.03°N, covering parts of Heilongjiang, Liaoning, Hebei, Shanxi, Henan, and Shaanxi provinces and the Inner Mongolia and Xinjiang Uyghur Autonomous Regions (Table 1). Each plant was collected at least 30 m away from its nearest neighbor.

2.2. Kernel oil content

Seeds were dried to a constant weight at 80°C and then pulverized in a ball mill. The kernel oil components were extracted with petroleum ether (boiling point 60°C) using a Soxhlet extraction device (Soxtec 8000, FOSS). The extraction process proceeded as follows: boiling at 120°C for 5 min, leaching for 1 h, and recovery for 25 min. The kernel oil weight was calculated by the weight difference between the sample extractions. Each sample was analyzed three times. Data are reported hereafter as mean ± standard deviation.

2.3. DNA extraction

Genomic DNA was extracted from young leaf tissues of each individual plant using TaKaRa MiniBEST Plant Genomic DNA Extraction Kit (Dalian, CN). After electrophoresis on 0.8% agarose gel, the DNA concentration was quantified using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA). The concentration of DNA was adjusted to 20 ng/μL, and the quality of the product was evaluated by determining both the 260/280-nm and 260/230-nm absorbance ratios, which should be >1.8 [24].

2.4. SSR markers and PCR amplification

We used 38 genomic, highly polymorphic SSR markers developed by Bi and Guan [23] to assess the genetic diversity in the 142 samples. PCR amplifications of all primers were performed in total volumes of 10 μL containing 20 ng DNA, 0.2 μM forward primer, 0.2 μM reverse primer, and 5 μL RR901A mix (Takara, Bio). PCR amplifications were performed in a thermal cycler (T100, BIO-RAD) in the following sequence: initial denaturation at 94°C for 3 min followed by 30 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 1 min. The final 10-min extension was performed at 72°C. The PCR products were first verified on 1.5% agarose gel. Subsequently, the

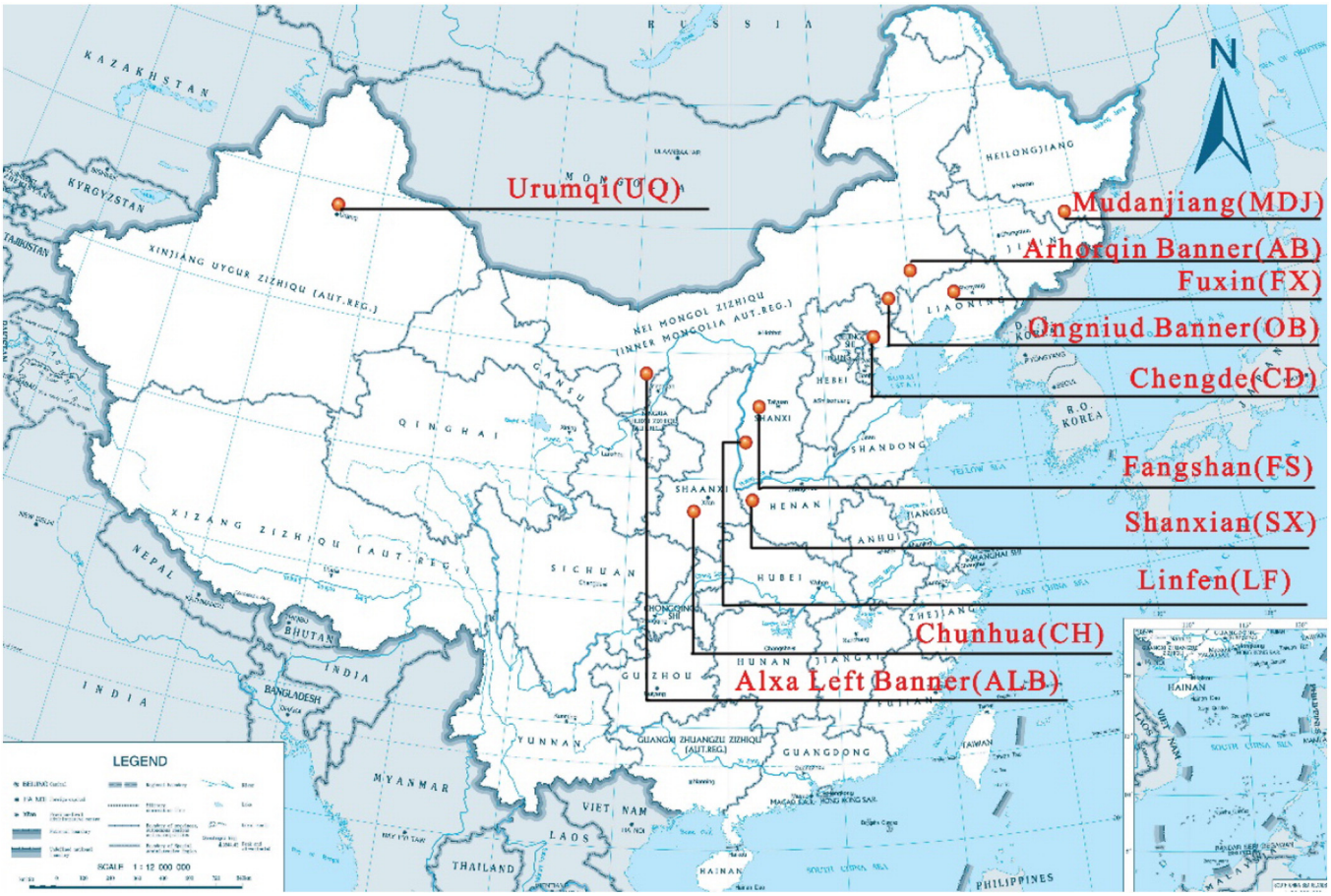

Fig. 1. Geographic locations (and codes) of 11 *X. sorbifolia* populations included in the study.

Table 1
Geographical and environmental parameters at 11 collecting sites.

Site code	Region	Longitude (N) (°)	Latitude (E) (°)	Elevation (m)	AMT (°C)	Jan MT (°C)	July MT (°C)	AR (mm)	ALH (h)	FFP (d)
MDJ	Heilongjiang	129.03	44.87	330	4.3	-17	22	540	2305	126
FX	Liaoning	122.00	42.17	630	7.6	-12.3	24	481	2735	154
AB	Inner Mongolia Autonomous Region	120.02	43.80	550	5.5	-13.3	24.7	365	2943.5	125
OB	Inner Mongolia Autonomous Region	119.10	42.37	700	5.8	-12.5	22.5	370	2925	115
ALB	Inner Mongolia Autonomous Region	105.68	38.85	1200	7.2	-8.5	23	150	3316	150
CD	Hebei	117.50	40.58	510	5.6	-9.4	24.4	536	2845	127
FS	Shanxi	111.38	38.04	1620	8.7	-8.2	22	520.1	2551.4	123
LF	Shanxi	110.70	36.60	1155	10.5	-2.5	26.5	500	2001.7	202
SX	Henan	111.62	34.63	1025.5	13.9	-0.7	27	650	2354.3	219
CH	Shaanxi	108.63	34.25	1301.2	9.8	-4.3	23.1	600.6	2372.7	183
UQ	Xinjiang Uygur Autonomous Region	87.97	43.76	622.5	6.9	-12.6	23.7	286.3	2813.5	181

AMT, annual mean temperature; Jan MT, January mean temperature; July MT, July mean temperature; AR, annual rainfall; ALH, annual daylight hours; FFP, frost-free period.
MDJ, Mudanjiang; FX, Fuxin; AB, Arhorqin Banner; OB, Ongniud Banner; ALB, Alxa Left Banner; CD, Chengde; FS, Fangshan; LF, Linfen; SX, Shanxian; CH, Chunhua; UQ, Urumqi.

primers with corresponding bands were resolved by nondenaturing polyacrylamide gel electrophoresis (PAGE) to check the DNA banding patterns and visualized by silver nitrate staining. The clear polymorphic bands were used in an identification step for the analysis of the genetic relationships and diversity among 11 populations. The reproducibility of PCR procedures was confirmed by repeating the process three times.

2.5. Data analysis

We examined the kernel oil contents by ANOVA. The distance matrices were based on the Gowers general similarity coefficient [25]. Cluster analyses using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) procedure were performed using NTSYS-pc 2.11 software (Exeter Software, Stauket, NY).

In our molecular analysis using SSR markers, the polymorphic bands were scored as either present (1) or absent (0). Alleles were alphabetically coded (e.g., A, B, or C for three bands) in decreasing size order. Samples with single bands were assumed to be homozygous. We calculated the number of alleles per locus (Na), number of effective alleles per locus (Ne), the Shannons information index (I), observed heterozygosity (Ho), and expected heterozygosity (He) using POPGENE 32 software [26]. The polymorphism information content (PIC) was estimated using Power Stats V12.xls software [27]. A matrix of genetic distances [28] was constructed for the 11 populations. A dendrogram cluster analysis was performed on NTSYS-pc 2.11 software using the UPGMA procedure. We investigated the distribution of genetic variation within and among populations by analysis of molecular variance (AMOVA) using GenAlEx 6.501 software; the procedure is based on the hierarchical variance of gene frequencies. A Mantel test was performed to examine the relationships among kernel oil content, genetic distance, and geographic distance among the 11 populations.

3. Results

3.1. Kernel oil content

Kernel oil contents of samples from the 11 populations are provided in Table 2. The oil content ranged from 43.53 ± 0.35% to 62.15 ± 18.92% among the locations, averaging 52.91 ± 4.73% (Table 2). The oil contents differed significantly among populations, and the pairwise differences between populations were frequently significant (Table 2).

The kernel oil content data were used to calculate genetic distances among populations, and the matrix of genetic distances was used to perform an agglomerative hierarchical cluster analysis (Fig. 2). At a distance coefficient of 5.08, the UPGMA dendrogram resolved the 11 populations into three major clusters. The first cluster was subdivided into three subclusters, one of which included four populations (MDJ, OB, LF, and CD). The second subcluster comprised populations ALB, SX, and UQ. The third subcluster contained population CH. The second cluster included populations FX and FS. The third cluster contained population AB.

3.2. Genetic diversity

We observed considerable variation in the amplified fragment patterns using different primers. Of the 38 primers tested, 15 yielded no amplification products; these are not included in our report. The remaining 23 SSR markers (Table 3) were used for the characterization and genetic diversity analysis of the 11 populations of *X. sorbifolia* (Table 4). We detected 80 alleles in total. Na values ranged from two (QXH002) to six (QXH274), averaging 3.48 alleles per locus across the 23 loci. All loci were polymorphic. PIC values ranged from 0.35 (QXH002) to 0.70 (QXH274), averaging 0.51. The mean He value was 0.58; values ranged from 0.45 in QXH002 to 0.75 in QXH274. Ho ranged from 41% in QXH177 to 92% in QXH365. Wright's fixation index (Fst) compares He and Ho, and is a measure of the degree of allelic fixation; Fst values ranged from 0.04 (in QBLB62 and QXR634) to 0.35 (in QXR343), averaging 0.13. The Shannon-Weaver information index ranged from 0.64 in QXH002 to 1.49 in QXH274, averaging 0.99. Thus, we detected abundant genetic diversity among the 11 populations. The most polymorphic locus, QXH274, had a high level of genetic variation. The high level of heterozygosity (average Ho = 0.73) detected by our SSR marker analysis indicated a high level of cross-pollination in *X. sorbifolia*.

Table 2
Kernel oil contents in *X. sorbifolia* populations.

Population	Kernel oil content (%)*
MDJ	53.60 ± 6.42^{abc}
FX	48.31 ± 0.84^{ab}
AB	62.15 ± 18.92^{c}
OB	53.81 ± 3.86^{bc}
ALB	51.62 ± 0.89^{abc}
CD	55.14 ± 5.35^{bc}
FS	43.53 ± 0.35^{a}
LF	53.84 ± 8.06^{bc}
SX	52.25 ± 6.48^{abc}
CH	56.92 ± 5.87^{bc}
UQ	50.88 ± 0.01^{abc}

Different lower case letters identify significantly different pairs of populations ($P < 0.05$; Duncan's post hoc test). Values are mean ± SD, n = 3.
MDJ, Mudanjiang; FX, Fuxin; AB, Arhorqin Banner; OB, Ongniud Banner; ALB, Alxa Left Banner; CD, Chengde; FS, Fangshan; LF, Linfen; SX, Shanxian; CH, Chunhua; UQ, Urumqi.
* $P < 0.05$ (ANOVA): kernel oil content differed significantly across the populations.

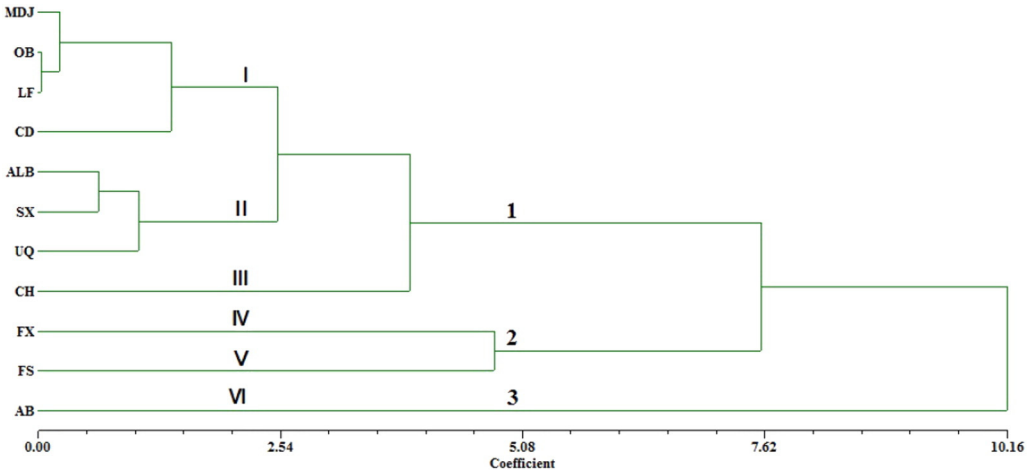

Fig. 2. Agglomerative hierarchical cluster analysis [using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) procedure] of the kernel oil contents of 11 *X. sorbifolia* populations. Gowers general similarity coefficients (Coefficient) were used to calculate genetic distances between pairs of populations. MDJ, Mudanjiang; FX, Fuxin; AB, Arhorqin Banner; OB, Ongniud Banner; ALB, Alxa Left Banner; CD, Chengde; FS, Fangshan; LF, Linfen; SX, Shanxian; CH, Chunhua; UQ, Urumqi.

Table 3
Characterization of 23 microsatellite loci in *X. sorbifolia.*

Locus	Repeat motifs	Primer sequences (5′–3′)	
QXH002	TC	F	AGAAGAACACTCAATGGGGA
		R	CTTCAACTGGACACCCGTAT
QXH049	CT	F	CCCCAACAAATGGTAAGACG
		R	GAATTTACAAGACAAGCAACAGC
QXH083	CT	F	AGCGGTCTCCTCCACTATCA
		R	GAATTGAAGCGCAGAAGGA
QXH120	AG	F	AAAACACTTCCGCACCAA
		R	TGGCTGCTGAGAAAGTAAGG
QXH177	AG	F	TGTGGTGGTTTTGGCAGAC
		R	CACCAAATAATGTCAATATCCTGT
QXH197	AG	F	GAAATATGAGGTCTTGGGTGTT
		R	GTGGCAGATAAACTGTCCTCAA
QXH262	CT	F	TCTAACCGAAGAAGCCAACT
		R	AGCGTGATATTCTGTTTCAACTAC
QXH274	CT	F	CATCGTCGTCTCATCCAGTAA
		R	GTGGCTTGTAGTTTGTTTCGTT
QXH282	CT	F	CCCAACAAATGGTAAGACGT
		R	GTTTCATTTCATTTCCAGCATC
QXH323	AG	F	CACAACCCAAATCCCAGAAC
		R	AACGACACGCACAATCATAAC
QXH365	AC	F	GTATATCTCTTTTACGGTCGTGAC
		R	ATGATGGGTTGGGTTGAGTT
QXHS371	GT	F	ATTGGAGTGGCCTTCATACG
		R	GCAAGCAGCTAAAGAAACAGC
QXH643	CA, CT	F	GCAGTTATGGAAAGGAATCA
		R	ATCAGTGTCGATTATTATCT
QXR343	TG	F	CACACTTTCTGAGTCCCGTAT
		R	TGTTTCTCCTCTAATCCAAC
QXR634	TTTTA, GA	F	CGTCCATCGCTTTCACCT
		R	AATAACAAATCAAAAATACCCCAT
QXR639	CT	F	CAACACCACCTCCACCAAC
		R	AAGGGATTTTGCTTTTCTGG
QXRB116	AC	F	CACCTTTCTCACTCCGTCTC
		R	CTCCTCGATCTGGTTGCTAA
QBRB203	AT	F	ACAAGTAGTGCTCATCGGTTTA
		R	GAGTCTAATAGGTAAGGCTAGGAAC
QBRS192	TC, CT	F	TTGTGCATCTGTGGAGAAGG
		R	TGCCTTCACATCCTGTGGTA
QBLB51	TC	F	TCCCACCCCAAAAAACTATAT
		R	CCTTCTGGAACTTGATGCC
QBLB58	TC, TA	F	CTCTTCCACTTCGGTCACG
		R	AACCAGAGCCACGCAAAT
QBLB62	CT	F	ATTCAGAAACGGTGGCTTAG
		R	TTATTTACTCTGGAAGGCTTATTT
QBLB65	TC	F	TATCTCGCCACATCTGCC
		R	CTGAAATGCCATTAACAAACAC

F, forward primer; R, reverse primer.

3.3. Populations structure

The mean proportion of polymorphic loci (P) within the population was 98.02%. *P*-values were 100% in nine of the populations, 86.96% in MDJ, and 91.30% in CH. Assuming the populations are in Hardy–Weinberg equilibrium, the mean genetic diversity (Nei) within populations was 0.50. Among the 11 populations, gene diversity ranged from 0.41 in MDJ to 0.57 in AB. The mean Shannons information index (I) was 0.79, ranging from 0.60 in MDJ to 0.93 in AB. The mean He was 0.53; values ranged from 0.48 in CH to 0.59 in FX. Ho ranged from 63% in CH to 80% in UQ. PIC values ranged from 0.37 in MDJ to 0.48 in AB, averaging 0.42 (Table 5). Nonhierarchical AMOVA analysis revealed that of the total SSR variation, 67% occurred among the populations, and the remainder within the populations.

Table 4
Genetic diversity among 11 populations of *X. sorbifolia.*

Locus	Na	Ne	I	Ho	He	Fst	PIC
QXH002	2	1.80	0.64	0.67	0.45	0.04	0.35
QXH049	4	2.33	1.03	0.79	0.57	0.08	0.51
QBLB51	4	2.52	1.09	0.89	0.61	0.17	0.55
QBLB58	3	2.37	0.93	0.85	0.58	0.06	0.49
QBLB62	3	1.96	0.76	0.76	0.49	0.04	0.39
QBLB65	3	2.33	0.93	0.74	0.57	0.16	0.48
QXH083	3	2.39	0.95	0.46	0.58	0.28	0.50
QXRB116	3	2.95	1.09	0.44	0.66	0.24	0.59
QXH120	3	2.36	0.93	0.75	0.58	0.10	0.48
QXH177	3	1.97	0.85	0.41	0.50	0.16	0.44
QBRS192	3	2.03	0.87	0.62	0.51	0.13	0.45
QXH197	3	2.31	0.92	0.62	0.57	0.22	0.48
QBRB203	3	2.47	1.00	0.88	0.60	0.15	0.53
QXH262	4	2.89	1.13	0.90	0.66	0.12	0.58
QXH274	6	3.89	1.49	0.91	0.75	0.20	0.70
QXH282	4	2.18	0.94	0.71	0.53	0.06	0.47
QXH323	4	3.19	1.24	0.88	0.69	0.08	0.63
QXR343	3	2.10	0.85	0.49	0.52	0.35	0.49
QXH365	3	2.97	1.09	0.92	0.67	0.07	0.59
QXHS371	3	2.66	1.03	0.89	0.63	0.15	0.54
QXR634	5	2.23	0.93	0.79	0.55	0.04	0.46
QXR639	5	2.28	1.06	0.63	0.56	0.07	0.51
QXH643	3	2.53	1.00	0.85	0.61	0.08	0.53
Mean	3.48	2.46	0.99	0.73	0.58	0.13	0.51

Na, number of alleles; Ne, number of effective alleles; Ho, observed heterozygosity; He, expected heterozygosity; I, Shannon's diversity index; Fst, fixation index; PIC, polymorphism information content.

Table 5
Genetic variability within 11 populations of *X. sorbifolia*.

Population	Sample size	(P) (%)	Ho	He	Nei	I	PIC
MDJ	6	86.96	0.68	0.49	0.41	0.60	0.37
FX	6	100.00	0.77	0.59	0.54	0.86	0.45
AB	40	100.00	0.73	0.57	0.57	0.93	0.48
OB	15	100.00	0.75	0.56	0.54	0.89	0.46
ALB	7	100.00	0.68	0.51	0.47	0.73	0.38
CD	15	100.00	0.76	0.52	0.50	0.79	0.41
FS	15	100.00	0.73	0.55	0.53	0.87	0.45
LF	16	100.00	0.66	0.52	0.50	0.83	0.43
SX	10	100.00	0.77	0.51	0.48	0.75	0.40
CH	6	91.30	0.63	0.48	0.43	0.68	0.39
UQ	6	100.00	0.80	0.56	0.52	0.80	0.42
Mean	12.55	98.02	0.72	0.53	0.50	0.79	0.42

P, percentage of polymorphic loci; Ho, observed heterozygosity; He, expected heterozygosity; I, Shannon's diversity index; Nei, genetic diversity; PIC, polymorphism information content. See Table 1 for population codes.
MDJ, Mudanjiang; FX, Fuxin; AB, Arhorqin Banner; OB, Ongniud Banner; ALB, Alxa Left Banner; CD, Chengde; FS, Fangshan; LF, Linfen; SX, Shanxian; CH, Chunhua; UQ, Urumqi.

3.4. Cluster analyses

To explore the genetic relationships of *X. sorbifolia* populations with different geographic provenance, we calculated Nei's genetic distances between pairs of populations. The matrix of genetic distances was then subjected to UPGMA cluster analysis (Fig. 3). The 11 populations were classified into three main groups with a coefficient value of 0.18. The first group was divided into two subgroups, one of which contained only population MDJ, whereas the other included five populations (FX, CD, AB, OB, and UQ). The second group was divided into two subgroups. The first contained only population FS; the second contained populations LF, SX and CH. The third group contained only population ALB.

A cluster analysis based on the similarity matrix of geographical provenances of the 11 populations separated the entire collection into three main groups with a coefficient level of 730.36 (Fig. 4). The first group was divided into two subgroups, one of which contained only population MDJ; the second subgroup included five populations (FX, CD, AB, OB, and UQ). The second group contained only population ALB. The third group was divided into two subgroups, one of which contained only population FS; the second subgroup contained populations LF, SX, and CH.

Mantel test analysis showed that the genetic and geographic distances of the 11 populations were significantly positively correlated ($r = 0.902$, $P < 0.01$) (Fig. 5).

4. Discussion

4.1. SSR markers

Data on genotypes can be used to identify the relationship between individuals and maintain genetic variation in germplasm, which is crucial for solving problems in breeding programs and for germplasm resource management [29]. Many types of molecular markers, especially SSR markers, have been used successfully to assess genetic diversity and characterize crop resources [30,31,32].

Molecular techniques based on DNA markers, such as RAPD and ISSR analysis, have been used to characterize the genetic diversity in *X. sorbifolia* [33,34,35], but SSRs markers have not yet come into general use in studies for this species. We used SSR analysis to investigate the genetic diversity in 11 *X. sorbifolia* populations; 23 polymorphic SSRs were highly informative. The proportion of polymorphic loci that we obtained (100.00%) exceeded the proportions in previous RAPD and ISSR studies [33,34,35]. The average number of alleles per locus was 3.48, and the average PIC value was 0.51. As previously demonstrated, the SSR assay approach is appropriate for genetic relationship studies [18,36], and it proved to be an efficient tool for the assessment of the genetic diversity of *X. sorbifolia* and identification of its populations in China.

4.2. Determinants of kernel oil content

The genetic diversity in kernel oil content was differentiated across the geographic range of *X. sorbifolia*. Kernel oil content varied among populations in different regions of China; the oil content of population AB was significantly higher than those of populations FX and FS. However, these phenotypic differences may have been the result of differences in environmental conditions across the country. Thus, we aimed to determine whether geographic provenance or genetic distance were responsible for variability in the kernel oil content. According to the cluster analyses (Fig. 2, Fig. 3, Fig. 4), the grouping of kernel oil content phenotypes was not concordant with measures of genetic distance or the geographic provenances of the populations. Using Mantel tests, we found no obvious relationship between

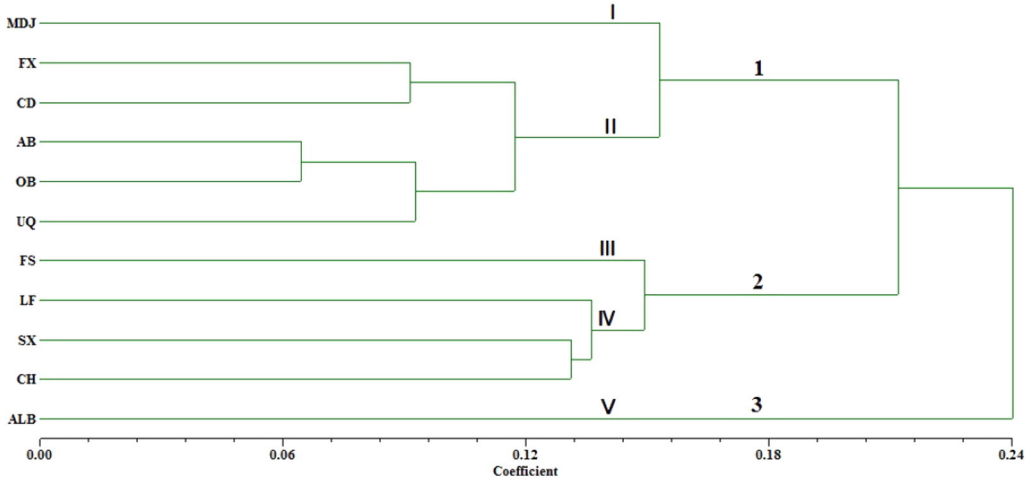

Fig. 3. Agglomerative hierarchical cluster analysis [using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) procedure] of Nei's pairwise genetic distances among 11 populations of *X. sorbifolia*. The dendrogram is based on the analyses of 23 simple sequence repeat loci. MDJ, Mudanjiang; FX, Fuxin; AB, Arhorqin Banner; OB, Ongniud Banner; ALB, Alxa Left Banner; CD, Chengde; FS, Fangshan; LF, Linfen; SX, Shanxian; CH, Chunhua; UQ, Urumqi.

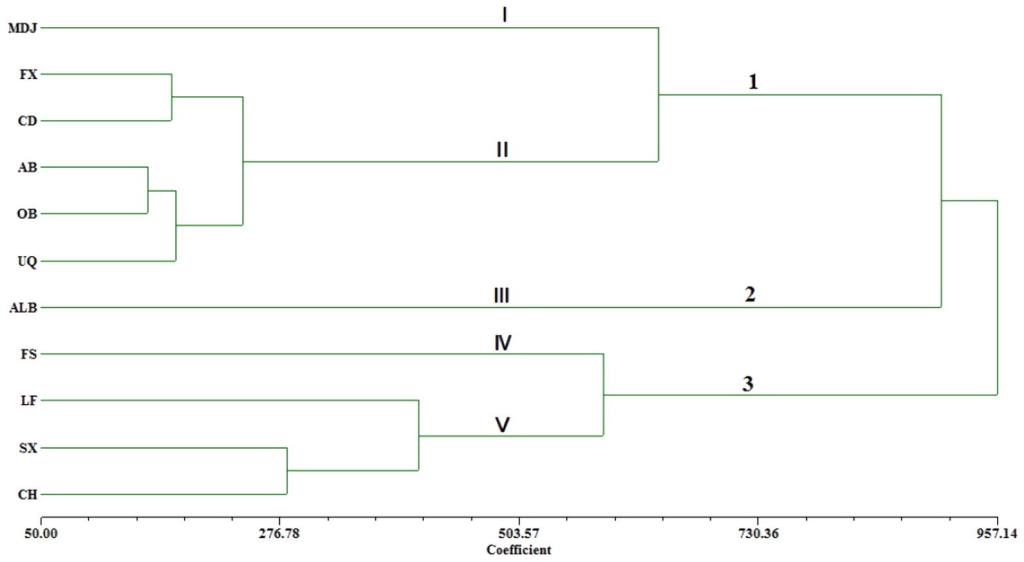

Fig. 4. Agglomerative hierarchical cluster analysis [using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) procedure] of Nei's pairwise genetic distances among 11 populations of *X. sorbifolia*. The dendrogram is based on the geographic provenances of the populations. MDJ, Mudanjiang; FX, Fuxin; AB, Arhorqin Banner; OB, Ongniud Banner; ALB, Alxa Left Banner; CD, Chengde; FS, Fangshan; LF, Linfen; SX, Shanxian; CH, Chunhua; UQ, Urumqi.

geographic provenance and kernel oil content or between genetic similarity and kernel oil content. Therefore, further studies are required to identify the determinants of kernel oil content variability among *X. sorbifolia* populations; these studies should include additional factors such as soil conditions and tree management protocols.

4.3. Levels of genetic variation within and among populations

Nybom [37] showed that mean genetic diversities within populations of self-pollinating, mixed-mating, and outcrossing species were 0.120, 0.180, and 0.250, respectively. We calculated a genetic diversity coefficient of 0.50 within populations of *X. sorbifolia*, a value that is markedly higher than those reported by Nybom [37]. We therefore concluded that *X. sorbifolia* had a cross-pollination mating system. The low diversity within populations may have been related to range contractions in recent decades due to human disturbances and natural causes [38]. With the low genetic diversity and fragmentation of wild *X. sorbifolia* populations, more stringent and efficient conservation policies should be established to maintain the genetic variation of the species as a resource for future breeding programs.

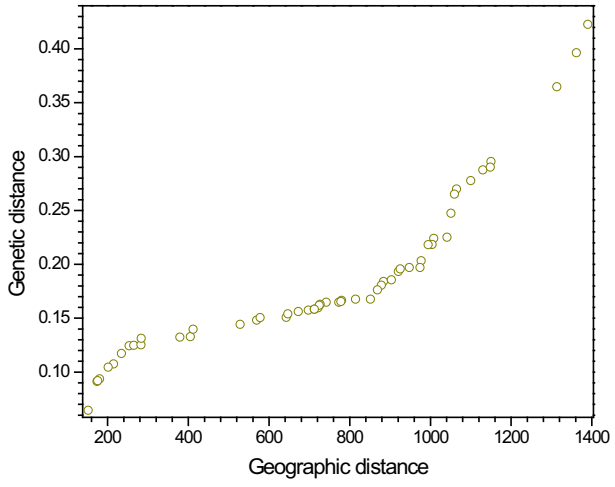

Fig. 5. Mantel test analysis of the correlation between geographic and genetic distances in 11 *X. sorbifolia* populations.

The 23 SSRs that were amplified were polymorphic in the 11 populations where large genetic diversity was observed. Between-population variation accounted for the largest proportion of genetic variability. The AMOVA indicated that between-population variation accounted for 67% of the total genetic variability, which identified a high degree of population differentiation in *X. sorbifolia*.

High degrees of population differentiation may be explained by several factors such as geographic isolation, habitat destruction, breeding system type, and restricted gene flow. Historical factors may also influence the distribution and partitioning of genetic diversity in plant species [39,40]. Our Mantel test analysis detected a significantly positive correlation between genetic and geographic distances among the 11 populations (Fig. 5), suggesting that these populations were adapted to local environmental conditions. Therefore, historical factors such as long-term isolation and habitat fragmentation may have influenced the geographic differentiation of *X. sorbifolia*. Gene flow among the 11 populations may be absent or limited because of physical barriers, such as high and large mountain ranges and deep and broad valleys. These barriers may have promoted genetic isolation and fragmentation among the populations [41]. Random losses of alleles may also have occurred during the times of geological transition and habitat fragmentation, and such losses would have contributed to population differentiation.

In addition to geographical factors, local-scale ecological processes such as seed dispersal, pollinator activities, breeding systems, and ongoing habitat destruction may have contributed to the genetic structure of *X. sorbifolia* stands. Our field observations indicated that the pollinators include bees, beetles, and syrphid flies, which have limited long-distance flight capabilities. Pollen transfer over long distances was therefore very unlikely as reported [40].

Thus, physical barriers, habitat fragmentation resulting from geological and climatic changes, and limited pollinator mobility may account for the observed low levels of within-population genetic diversity and the high levels of between-population genetic differentiation. The parallel and convergent evolution of *X. sorbifolia* populations should provide an interesting case study for exploring the mechanisms of diversification and adaptation in plants.

Our study contributed important information to improve the germplasm of *X. sorbifolia* for higher and better oil yields through the breeding of geographically isolated populations. Our data will

contribute to the management planning for the conservation of this important plant genetic resource.

5. Conclusion

We found significant variation in kernel oil content and microsatellite DNA polymorphisms among 11 populations of *X. sorbifolia*. There was no obvious relationship between geographical provenance and kernel oil content or between genetic similarity and kernel oil content, suggesting that attempts to improve the oil content of this species should focus on other factors. The large average number of alleles per locus and the allelic diversity in the set of genotypes analyzed indicate that the genetic spectrum in the material was relatively wide. The allelic abundance suggests the presence of potentially unique and rare alleles. The significantly positive correlation between genetic and geographic distances suggests adaptation to local conditions. SSR markers proved to be a useful tool for exploring plant diversity and will substantially contribute to progress in *X. sorbifolia* breeding programs.

Conflict of interest

The authors declare no conflict of interest.

Financial support

The study was financially supported by the International S&T Cooperation Program of China (2014DFA31140).

Acknowledgements

We thank Li Wang (College of Forestry, Beijing Forestry University) for assistance in revising the manuscript.

References

[1] Sheng DC, Feng QR, Li FL, Liu ZH. Pollination characteristics and fruit developing law of *Xanthoceras sorbifolia*. Liaoning For Sci Technol 2010;4:1–4.
[2] Li J, Fu YJ, Qu XJ, Wang W, Luo M, Zhao CJ, et al. Biodiesel production from yellow horn (*Xanthoceras sorbifolia* Bunge.) seed oil using ion exchange resin as heterogeneous catalyst. Bioresour Technol 2012;108:112–8. http://dx.doi.org/10.1016/j.biortech.2011.12.129.
[3] Yao ZY, Qi JH, Yin LM. Biodiesel production from *Xanthoceras sorbifolia* in China: Opportunities and challenges. Renew Sustain Energy Rev 2013;24:57–65. http://dx.doi.org/10.1016/j.rser.2013.03.047.
[4] Ling JH, Liu LL, Wang YX, Li ZY, Liu R, Li Q, et al. Characterization and quantification of the triterpenoids in different parts of *Xanthoceras sorbifolia* by HPLC–ESI–MS. J Pharm Biomed Anal 2011;55:259–64. http://dx.doi.org/10.1016/j.jpba.2011.01.030.
[5] Li ZL, Li DY, He XM, Hua HM. Two new triterpenoid saponins from the husks of *Xanthoceras sorbifolia*. Nat Prod Res 2013;27:232–7. http://dx.doi.org/10.1080/14786419.2012.671316.
[6] Xiao W, Wang Y, Zhang P, Li N, Jiang S, Wang JH, et al. Bioactive barrigenol type triterpenoids from the leaves of *Xanthoceras sorbifolia* Bunge. Eur J Med Chem 2013;60:263–70. http://dx.doi.org/10.1016/j.ejmech.2012.12.022.
[7] Zhang Q, Li J, Zhao YB, Korban SS, Han YP. Evaluation of genetic diversity in Chinese wild apple species along with apple cultivars using SSR markers. Plant Mol Biol Report 2012;30:539–46. http://dx.doi.org/10.1007/s11105-011-0366-6.
[8] Sreekumar VB, Renuka C. Assessment of genetic diversity in *Calamus thwaitesii* BECC. (Arecaceae) using RAPD markers. Biochem Syst Ecol 2006;34:397–405. http://dx.doi.org/10.1016/j.bse.2005.12.002.
[9] Ergül A, Tuerkoglu M, Soylemezoglu G. Genetic identification of Amasya (*Vitis vinifera* L. Cvs.) genotypes based on AFLP markers. Biotechnol Biotechnol Equip 2004;18:39–43. http://dx.doi.org/10.1080/13102818.2004.10817118.
[10] Mahar KS, Rana TS, Ranade SA. Molecular analyses of genetic variability in soap nut (*Sapindus mukorossi* Gaertn). Ind Crop Prod 2011;34:1111–8. http://dx.doi.org/10.1016/j.indcrop.2011.03.029.
[11] Sabir JS, Abo-Aba S, Bafeel S, Zari TA, Edris S, Shokry AM, et al. Characterization of ten date palm (*Phoenix dactylifera* L.) cultivars from Saudi Arabia using AFLP and ISSR markers. C R Biol 2014;337:6–18. http://dx.doi.org/10.1016/j.crvi.2013.11.003.
[12] Yang M, Han YN, Xu LM, Zhao JR, Liu YL. Comparative analysis of genetic diversity of lotus (*Nelumbo*) using SSR and SRAP markers. Sci Hortic 2012;142:185–95. http://dx.doi.org/10.1016/j.scienta.2012.05.021.
[13] Weber JL, Wong C. Mutation of human short tandem repeats. Hum Mol Genet 1993;2:1123–8. http://dx.doi.org/10.1093/hmg/2.8.1123.
[14] Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, et al. A gene map of the human genome. Science 1996;274:540–6. http://dx.doi.org/10.1126/science.274.5287.540.
[15] György Z, Vouillamoz JF, Höhn M. Microsatellite markers reveal common East Alpine–Carpathian gene pool for the arctic-alpine *Rhodiola rosea* (Crassulaceae). Plant Syst Evol 2016;302:721–30.
[16] Tanya P, Taeprayoon P, Hadkam Y, Srinives P. Genetic diversity among Jatropha and Jatropha-related species based on ISSR markers. Plant Mol Biol Report 2011;29:252–64. http://dx.doi.org/10.1007/s11105-010-0220-2.
[17] Pamidimarri DVNS, Singh S, Mastan SG, Patel J, Reddy MP. Molecular characterization and identification of markers for toxic and non-toxic varieties of *Jatropha curcas* L. using RAPD, AFLP and SSR markers. Mol Biol Rep 2009;36:1357–64. http://dx.doi.org/10.1007/s11033-008-9320-6.
[18] Ben-Ayed R, Sans-Grout C, Moreau F, Grati-Kamoun N, Rebai A. Genetic similarity among Tunisian olive cultivars and two unknown feral olive trees estimated through SSR markers. Biochem Genet 2014;52:258–68. http://dx.doi.org/10.1007/s10528-014-9645-x.
[19] Xu M, Sun YG, Li H. EST-SSRs development and paternity analysis for *Liriodendron* spp. New For 2010;40:361–82. http://dx.doi.org/10.1007/s11056-010-9205-0.
[20] Yang AH, Zhang JJ, Tian H, Yao XH. Characterization of 39 novel EST-SSR markers for *Liriodendron tulipifera* and cross-species amplification in *L. Chinese* (Magnoliaceae). Am J Bot 2012;99:460–4. http://dx.doi.org/10.3732/ajb.1200154.
[21] Pelsy F, Hocquigny S, Moncada X, Barbeau G, Forget D, Hinrichsen P, et al. An extensive study of the genetic diversity within seven French wine grape variety collections. Theor Appl Genet 2010;120:1219–31. http://dx.doi.org/10.1007/s00122-009-1250-8.
[22] Riahi L, Zoghlami N, El-Heit K, Laucou V, Le Cunff L, Boursiquot J, et al. Genetic structure and differentiation among grapevines (*Vitis vinifera*) accessions from Maghreb region. Genet Resour Crop Evol 2010;57:255–72. http://dx.doi.org/10.1007/s10722-009-9467-6.
[23] Bi QX, Guan WB. Isolation and characterisation of polymorphic genomic SSRs markers for the endangered tree *Xanthoceras sorbifolium* Bunge. Conserv Genet Resour 2014;6:895–8. http://dx.doi.org/10.1007/s12686-014-0234-4.
[24] Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: A laboratory manual. 2nd ed. NY, USA: Cold Spring Harbor Laboratory Press, Cold Spring Harbor; 1989.
[25] Gower JC. A general coefficient of similarity and some of its properties. Biometrics 1971;27:857–71. http://dx.doi.org/10.2307/2528823.
[26] Yeh F, Boyle T. Population genetic analysis of co-dominant and dominant markers and quantitative traits. Belg J Bot 1997;129:157.
[27] Brenner C, Morris JW. Paternity index calculations in single locus hypervariable DNA probes: Validation and other studies. Proceedings for the International Symposiumon Human Identification. Madison, USA: Promega Corporation; 1990. p. 21–53.
[28] Nei M. Genetic distance between populations. Am Nat 1972;106:283–91. http://dx.doi.org/10.1086/282771.
[29] Salem KFM, El-Zanaty AM, Esmail RM. Assessing wheat (*Triticum aestivum* L.) genetic diversity using morphological characters and microsatellite markers. Word J Agric Sci 2008;5:538–44.
[30] Keneni G, Bekele E, Imtiaz M, Dagne K, Getu E, Assefa F. Genetic diversity and population structure of Ethiopian chickpea (*Cicer arietinum* L.) germplasm accessions from different geographical origins as revealed by microsatellite markers. Plant Mol Biol Report 2012;30:654–65. http://dx.doi.org/10.1007/s11105-011-0374-6.
[31] McClean PE, Terpstra J, McConnell M, White C, Lee R, Mamidi S. Population structure and genetic differentiation among the USDA common bean (*Phaseolus vulgaris* L.) core collection. Genet Resour Crop Evol 2012;59:499–515. http://dx.doi.org/10.1007/s10722-011-9699-0.
[32] Suresh S, Park JH, Cho GT, Lee HS, Baek HJ, Lee SY, et al. Development and molecular characterization of 55 novel polymorphic cDNA-SSR markers in faba bean (*Vicia faba* L.) using 454 pyrosequencing. Molecules 2013;18:1844–56. http://dx.doi.org/10.3390/molecules18021844.
[33] Guan LP, Yang T, Li N, Li BS, Lu H. Identification of superior clones by RAPD technology in *Xanthoceras sorbifolia* Bge. For Stud China 2010;12:37–40. http://dx.doi.org/10.1007/s11632-010-0003-8.
[34] Zhang YX, Liu JJ, Bai JH, Guo HY, Guo JP. Optimization of ISSR-PCR analysis for *Xanthoceras sorbifolia* by uniform design application. J Shanxi Agric Univ 2014;34:241–8.
[35] Lu J, Chai CS, Wu WJ, Qi JL. Optimization of ISSR-PCR amplification in *Xanthoceras sorbifolia* Bunge. Chin Agric Sci Bull 2014;30:32–6.
[36] Sonnante G, Carluccio AV, Paolis AD, Pignone D. Identification of artichoke SSR markers: Molecular variation and patterns of diversity in genetically cohesive taxa and wild allies. Genet Resour Crop Evol 2008;55:1029–46. http://dx.doi.org/10.1007/s10722-008-9310-5.
[37] Nybom H. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol Ecol 2004;13:1143–55. http://dx.doi.org/10.1111/j.1365-294X.2004.02141.x.
[38] Zhang XM, Wen J, Dao ZL, Motley TJ, Long CL. Genetic variation and conservation assessment of Chinese populations of *Magnolia cathcartii* (Magnoliaceae), a rare evergreen tree from the South-Central China hotspot in the Eastern Himalayas. J Plant Res 2010;123:321–31. http://dx.doi.org/10.1007/s10265-009-0278-9.
[39] Parks CR, Wendel JF, Swell MM, Qiu YL. The significance of allozyme variation and introgression in the *Liriodendron tulipifera* complex (Magnoliaceae). Am J Bot 1994;81:878–89. http://dx.doi.org/10.2307/2445769.
[40] Kamm U, Rotach P, Gugerli F, Siroky M, Edwards P, Holderegger R. Frequent long-distance gene flow in a rare temperate forest tree (*Sorbus domestica*) at the landscape scale. Heredity 2009;103:476–82. http://dx.doi.org/10.1038/hdy.2009.70.
[41] Matsuki Y, Tateno R, Shibata M, Isagi Y. Pollination efficiencies of flower-visiting insects as determined by direct genetic analysis of pollen origin. Am J Bot 2008;95:925–30. http://dx.doi.org/10.3732/ajb.0800036.