Genetics and Molecular Research

Genetic diversity in fragmented populations of *Populus talassica* inferred from microsatellites: implications for conservation

X.H. Zhu[1,3], S.P. Cheng[2], T. Liao[1] and X.Y. Kang[1]

[1]National Engineering Laboratory for Tree Breeding,
Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants,
Ministry of Education, Beijing Laboratory of Urban and Rural Ecological
Environment, College of Biological Sciences and Technology,
Beijing Forestry University, Beijing, China
[2]Pingdingshan University, Pingdingshan, Henan Province, China
[3]College of Forestry and Horticulture,
Xinjiang Agricultural University, East Nongda Road, Urumqi, China

Corresponding author: X.Y. Kang
E-mail: kangxy@bjfu.edu.cn

Genet. Mol. Res. 15 (2): gmr.15027899
Received October 22, 2015
Accepted January 18, 2016
Published May 25, 2016
DOI http://dx.doi.org/10.4238/gmr.15027899

ABSTRACT. *Populus talassica* Kom. is an ecologically important species endemic to central Asia. In China, its main distribution is restricted to the Ili region in the Xinjiang Autonomous Region. An understanding of genetic diversity and population structure is crucial for the development of a feasible conservation strategy. Twenty-six high-level simple sequence repeat (SSR) markers were screened and used to genotype 220 individuals from three native populations. A high level of genetic diversity and low population differentiation were revealed. We identified 163 alleles, with a mean of 6.269 alleles per locus. The observed and expected heterozygosities ranged from 0.472 to 0.485 (with a mean of 0.477), and from 0.548 to 0.591 (mean 0.569), respectively. Analysis of molecular variance revealed 93% variation within populations and

 ©FUNPEC-RP www.funpecrp.com.br

7% among populations. A model-based population structure analysis divided *P. talassica* into two groups (optimal K = 2). These genetic data provide crucial insight for conservation management.

Key words: *Populus talassica*; Genetic diversity; Population structure; Microsatellite

INTRODUCTION

Populus talassica Kom., an outcrossing perennial species, belongs to the section *Tacamahaca Spach* of the genus *Populus*. It is native to central Asia, with fragmented and isolated populations. In China, its main distribution is from the center to the west of the Tian Shan Mountains in the Xinjiang Autonomous Region (Flora of China, http://www.efloras.org/florataxon.aspx?flora_id=2&taxon_id=200005721). It plays a major role in ecological and environmental protection. In this area, the largest and the best-preserved wild population is spread along the Kashi river basin in the Ili region. Throughout the valley, on both sides of the river (about 110 Km), individuals are sporadically distributed in an area of approximately 1400 ha. However, there are three relatively intensive areas, the upper, middle, and lower reaches of the river, respectively. This area was listed as the "Key National Charity Protecting Forests" in 2005 (http://www.iyaxin.com).

P. talassica plays a significant role in flood control, water conservation, soil erosion prevention, biodiversity maintenance, and so on. However, the wild population has declined sharply because of deforestation, over-grazing, and improper management (Personal observation). People have little knowledge of this species and underestimate its importance to local and regional ecology. Some localities were deforested, and many individuals were secretly felled for firewood. Due to lack of management, mismanagement, and various other reasons, forests have severely declined, and the ecological benefits of *P. talassica* are becoming increasingly vulnerable.

A large body of research has determined that a decline in population size could lead to the increased probability of extinction (Barrett and Kohn, 1991; Lande, 1993, 1994; Newman and Pilson, 1997; Higgins and Lynch, 2001) and accordingly work has been devoted to particular issues in order to optimize conservation management (Newman and Pilson, 1997; Hedrick and Kalinowski, 2000; Newman and Tallmon, 2001; Keller and Waller, 2002; Hufford and Mazer, 2003; Tallmon et al., 2004; Volis and Blecher, 2010). To date, however, few studies have investigated the genetic germplasm of wild *P. talassica.* Previous studies were mainly focused on salt tolerance and cold resistance when crossing parents (Bo, 2003; Zhang et al., 2010). In the assessment of genetic variation information, microsatellites (simple sequence repeats [SSRs]) are ideal because of the high levels of polymorphism, co-dominance, reproducibility, and transferability (Tóth et al., 2000). This genetic information has been widely shown to be essential for informing conservation management decision-making.

A large number of SSRs have been developed for species in the genus *Populus*, including *P. trichocarpa*, *P. tremuloides*, *P. tomentosa*, and *P. szechuanica* (Rahman et al., 2000; Yin et al., 2009; Du et al., 2012b; Shen et al., 2014). To design feasible conservation strategies, it is essential to develop species-specific SSR markers that can be used to assess genetic diversity and population structure in *P. talassica*.

 ©FUNPEC-RP www.funpecrp.com.br

In this paper, 920 SSR primers were downloaded from the *Populus* Molecular Genetics Cooperative (GCPM and PMGC primers; http://www.ornl.gov) and the Oak Ridge National Laboratory (ORPM primers; Tuskan et al., 2004). After screening, 26 markers exhibited robust amplification and high levels of polymorphism. These markers were then used to evaluate genetic variation within and among populations. These genetic data are an indispensable foundation for understanding the genetic basis of wild *P. talassica* populations and facilitating future conservation management.

MATERIAL AND METHODS

Sampling strategy and DNA extraction

According to our sampling scheme, three native populations were sampled from the upper, middle, and lower reaches of the Kashi River, as described in the Introduction. These three geographically separate sites were tens of kilometers apart (Figure 1). We selected individuals at least 30 m apart to prevent selection of clones. Finally, a total of 220 individuals were sampled, and the location of each tree was determined using a Garmin (Xinbei, Taiwan) global positioning system device (Table 1). In the field survey, we also made a biological characteristic statistic for each individual, including age (depending on growth cone), diameter at breast height and height.

Figure 1. *Populus talassica* population locations. In the Kashi river basin, we selected three geographically separate sites to select samples.

Table 1. Locations of native *Populus talassica* populations used in this study.

Population code	Sampling localities	Latitude (N)	Longitude (E)	Altitude (m)	Sample size
Upper	Kemeng County, Xinjiang	43°76′	82°63′	1150	66
Middle	Keling County, Xinjiang	43°80′	82°44′	1050	75
Lower	Qiahawulasitai, Xinjiang	43°83′	82°24′	968	79

The branches were cultured in water using a typical “twig and pot water culture” in a greenhouse to force floral development (Dong et al., 2014). Total genomic DNA was extracted from approximately 2 g leaf tissue using a DNeasy Plant Mini Kit (Tiangen Biotech Co. Ltd., Beijing, China) following the manufacturer’s instructions. The quality and concentration of

the extracted DNA were determined using 1% agarose gel electrophoresis and ultraviolet spectrophotometry. The DNA samples were diluted to 20-50 ng/μL for use as the template for polymerase chain reaction (PCR) amplification.

Primer selection and amplification

A total of 920 SSR primers were downloaded from the *Populus* Molecular Genetics Cooperative (http://www.ornl.gov; GCPM and PMGC primers) and the Oak Ridge National Laboratory (ORPM primers; Tuskan et al., 2004). DNA extracted from two individuals was amplified to test the transferability and suitability of SSRs in *P. talassica.* Eight individuals were then used to test for polymorphisms in markers that exhibited robust amplification.

The forward primer of each pair was tagged with the universal M13 sequence (5'-TGTAAAACGACGGCCAGT-3') during synthesis. Each PCR used a 20-μL total volume containing 10 μL 2X Tap PCR Mix (Biomedtech, Beijing, China), 0.4 μL fluorescent-dye-labeled (FAM, HEX, TAMRA, ROX) M13 primer, 0.08 μL forward primer, 0.32 μL reverse primer, 7.2 μL ddH_2O, and 2 μL (~20 ng) genomic DNA. PCR amplifications were performed in a GeneAmp® PCR System 9700 thermal cycler (Bio-Rad Laboratories Inc., Beijing, China) using the following program: 94°C for 10 min; followed by 30 cycles of 30 s at 94°C, 40 s at 53°C, and 40 s at 72°C; then 10 cycles of 30 s at 94°C, 40 s at 50°C, and 40 s at 72°C; and a final extension at 72°C for 10 min. The PCR products were resolved using an ABI 3730XL DNA Analyzer by Genewiz Biotechnology Co., Ltd., (Beijing, China), and the data were analyzed using the Gene-Marker software (SoftGenetics LLC, State College, PA, USA).

Data analysis

Possible null alleles and genotyping errors caused by stuttering and/or large-allele dropout were tested using MICRO-CHECKER (1000 randomizations; Van Oosterhout et al., 2004). For samples with poor amplification, a duplicate reaction was conducted. The estimated parameters of genetic diversity comprised the number of alleles (N_A), effective number of alleles (N_E), observed (H_O) and expected (H_E) heterozygosities, inbreeding coefficients (F_{IS}), and pair-wise differentiation among subpopulations (F_{ST}), which were assessed using GeneAlEx 6.2 (Peakall and Smouse, 2006). Polymorphic information content (PIC) was calculated using CERVUS 3.0 (Marshall et al., 1998).

The Hardy-Weinberg equilibrium (HWE) expectations and genotypic linkage disequilibriums (LDs) between all pairs of loci were implemented in Genepop v. 4.0 (Raymond and Rousset 1995; Rousset 1997).

To assess the spatial genetic structure, we used a Bayesian analysis implemented in STRUCTURE (Pritchard et al., 2000) to examine the number of differentiated populations. The K was set from 2 to 8 with each K estimate replicated 10 times with a 100,000-burn-in period and 100,000 Markov chain Monte Carlo (MCMC) iterations. The optimum number of clusters was calculated by uploading results to Structure Harvester (http://taylor0.biology.ucla.edu/structureHarvester/) (Pritchard et al., 2000; Evanno et al., 2005; Jakobsson and Rosenberg, 2007; Earl and vonHoldt 2012).

Analysis of molecular variance (AMOVA) was performed to partition the genetic variance, and a Mantel test was carried out to associate genetic and geographical distance using GeneAlEx 6.2 (Peakall and Smouse, 2006).

RESULTS

Genetic diversity

SSRs are widely used in genetic diversity studies as neutral markers. In this study, of 920 SSR primers, 126 markers were successfully amplified in *P. talassica*, and 26 markers exhibited at least two alleles in a panel of eight individuals. After checking with Micro-Checker, no null alleles were found in any loci. The PIC value, suggesting the polymorphic level of the loci, was 0.544 ± 0.212, and ranged from 0.007 for GCPM_2613-1 to 0.867 for GCPM_672-1. Of the 26 loci, 13 had PIC values above 0.5 (Table 2). No significant LD was detected between any of the SSR loci pairs (data not shown), which suggests that these 26 SSRs could be used as independent markers to assess genetic diversity and population structure in *P. talassica.*

Table 2. Diversity information parameters of 26 SSR markers at *Populus talassica* species level.

Locus	Forward primers	Reverse primers	Motif	Length	N_A	N_E	PIC	H_O	H_E	*F*
GCPM_3205	TCCTAACCCGTGACTCTCTA	GTGGTGTTGGTTTTGAGTCT	(GAA)8	179	7.000	4.333	0.734	0.402	0.769	0.477
GCPM_1781-1	AACCAAAGCATCAAGCATAG	AGACATGGCGTAATCACTCT	(AT)14	135	3.000	2.982	0.591	0.623	0.665	0.063
GCPM_2216-1	GGGATTCCTCTGCTCTAAAT	TTGATCATCAGAGGCCTAAT	(TA)13	200	11.000	7.354	0.849	0.464	0.864	0.463
ORPM_369	TGTTCGGGTTATATTGCCATT	TGATTGGGTGTCTCTGCTTG	(ATA)7	203	2.000	1.306	0.207	0.271	0.234	-0.156
GCPM_3869-1	GAAATATTTTGCCATAGTTATTATT	AAAACAACCACAAGAAATCC	(AT)13	165	10.000	4.524	0.760	0.313	0.779	0.598
GCPM_3677-1	ATGCTCGAGGAAATATCAAA	TGCATCAGAGATCAGAATGA	(TA)22	208	10.000	3.970	0.721	0.656	0.748	0.124
GCPM_2127-1	GGCAGTTGAAATATTCGAGA	CGAGTCTTCTTTCGAACCTA	(AG)10	160	3.000	1.800	0.348	0.514	0.444	-0.156
GCPM_672-1	GAAAGAAAATGAACCCATCA	CCAGGATATAATGTCCCAGA	(GT)13	197	14.000	8.230	0.867	0.769	0.878	0.125
GCPM_3390-1	CAGAGCTCTAACCATGGAAC	GCTACGGTGATGGTTTGTAT	(TA)19	220	12.000	2.453	0.577	0.343	0.592	0.421
GCPM_1838	GTTCAGCGAAAGCTAAAGAG	CACAGAATTACAGCTGATGC	(GA)14	140	13.000	5.305	0.787	0.845	0.812	-0.042
GCPM_4045-1	TGCCATATGCATTAATCTTG	TAAGCAAAGTCGATAACCGT	(AT)33	207	4.000	2.209	0.490	0.298	0.547	0.456
GCPM_4046-1	TTCCAAAGGTAAGGTGGTTA	GAGGTTGCTTTCTCTTGCTA	(CT)18	153	14.000	6.206	0.822	0.745	0.839	0.112
ORPM_162	TGTCGAGAAGTAATATTGCACCA	GAAGCATATGCAGGGCAAC	(GT)4	213	2.000	1.944	0.368	0.159	0.485	0.672
GCPM_3586	TGGAGTTTCCTCATCTCTTG	CGCTTCAAGTAAGTGTGTGT	(TC)9	225	4.000	2.466	0.533	0.190	0.594	0.680
GCPM_1676-1	CAACAAAATCAATCGCTCTC	ACCCTAGCAAAATCAACAAA	(TC)9	134	8.000	4.153	0.725	0.764	0.759	-0.006
GCPM_2613-1	TACACACCATGGGGATTTAT	CTGTGACCCGATCTTTTAGA	(AT)24	167	2.000	1.088	0.077	0.056	0.081	0.304
GCPM_3264-1	GTCGGTTTCATACCCAATAA	ATTCGGATAATGTGAAGTGC	(CT)11	197	2.000	1.872	0.357	0.382	0.466	0.180
GCPM_3937-1	ATTGAAAATGTTGCAAGTCG	ATGAGAATTGAGATGCCAAC	(CA)10	166	5.000	1.617	0.348	0.409	0.382	-0.072
PMGC_2522	TCTGTTAATTTCTCAGCTGTTG	TGCTTTACTAAACTTTTTACTGC	(GA)	175	5.000	3.664	0.689	0.654	0.727	0.100
GCPM_1373-1	AAATCCCACTTCCGTTAAAT	AAAAGTTATTTGCTTGCTGC	(CT)9	225	3.000	1.771	0.345	0.484	0.435	-0.111
GCPM_2791-1	CTTCGAGCTTCTCAAAAAGA	GGAGATTCTGACGAGGGT	(AT)17	192	4.000	1.710	0.382	0.324	0.415	0.220
GCPM_4029-1	ATTTGGATTTCCATGTTCAG	CATGAGTTAGCAACGAAACA	(CA)10	172	7.000	4.405	0.737	0.815	0.773	-0.054
GCPM_2705-1	TTTGCACAGGTAAGTTGATG	ATTGACCATAGCAGACAACC	(AC)9	216	5.000	2.213	0.499	0.564	0.548	-0.028
PMGC_2558	CCAGAGAAAGAGAGTGCTTC	AATGCAGATGTCGTTGTTTGC	(GA)	155	5.000	2.212	0.480	0.488	0.548	0.109
GCPM_2615-1	ATGTCAACGTCACTGACAAA	ATTAGGCAATGCAGAACACT	(CTT)5	222	4.000	1.918	0.384	0.427	0.479	0.108
ORPM_136	TTTAAGCCTCCGAAAACCAA	TTTAAGCCTCCGAAAACCAA	(CT)6	209	4.000	1.998	0.454	0.414	0.499	0.171
Mean	-	-	-	-	6.269	3.219	0.544	0.476	0.591	0.183
St. Dev	-	-	-	-	0.774	0.371	0.212	0.041	0.039	0.049

N_A = number of alleles per locus; N_E = effective number of alleles; PIC = polymorphism information content; H_O = observed heterozygosity; H_E = expected heterozygosity; *F* = Wright's fixation index.

In total, 163 alleles were identified, with a mean of 6.269 ± 0.774 alleles per locus. Among them, locus GCPM_672-1 was observed with 14 alleles, exhibiting the greatest variation. N_E ranged from 8.230 for GCPM_672-1 to 1.088 for GCPM_2613-1, with a mean of 3.219. H_O and H_E were meaningful parameters for accessing genetic diversity. H_O and H_E ranged from 0.472 to 0.485 (mean 0.477), and from 0.548 to 0.591 (mean = 0.569), respectively (Table 2).

At the population level, almost all genetic parameters were similar, indicating that genetic variation was not significant among the three geographical populations. N_a and N_e ranged from 4.885 (Upper population) to 6.000 (Lower) and from 2.908 (Middle) to 3.288 (Lower), respectively. The highest H_O was 0.485 in the Middle population, followed by 0.474

in the Upper population, and 0.472 in the Lower population. However, H_E differed from H_O. The highest was in the Lower population (0.591), and the lowest was in the Middle population (0.567). The allele equilibrium statuses of the three populations were tested for HWE, and 23 of 78 locus-population comparisons, deviated significantly from HWE ($P < 0.05$) (Table 3).

Table 3. Comparisons of genetic diversity and differentiation at *Populus talassica* population level.

Locus	Upper					Middle					lower					All Pops			
	N_A	N_E	H_O	H_E	F	N_A	N_E	H_O	H_E	F	N_A	N_E	H_O	H_E	F	F_{IS}	F_{IT}	F_{ST}	N_M
GCPM_3205	5.000	4.608	0.700	0.783	0.106	6.000	2.993	0.151	0.666	0.774	7.000	4.204	0.447	0.762	0.413	0.413	0.442	0.049	4.850
GCPM_1781-1	3.000	2.431	0.754	0.589	-0.281	3.000	2.989	0.667	0.665	-0.002	3.000	2.920	0.481	0.658	0.268	0.005	0.044	0.039	6.129
GCPM_2216-1	10.000	6.070	0.585	0.835	0.300	9.000	6.397	0.500	0.844	0.407	8.000	5.869	0.324	0.830	0.609	0.438	0.457	0.032	7.449
ORPM_369	2.000	1.238	0.215	0.192	-0.121	2.000	1.557	0.467	0.358	-0.304	2.000	1.136	0.128	0.120	-0.068	-0.209	-0.156	0.044	5.423
GCPM_3869-1	8.000	6.230	0.435	0.839	0.481	7.000	2.199	0.227	0.545	0.584	10.000	5.518	0.299	0.819	0.635	0.564	0.592	0.065	3.625
GCPM_3677-1	8.000	2.929	0.723	0.659	-0.098	8.000	3.178	0.757	0.685	-0.104	10.000	5.597	0.486	0.821	0.409	0.092	0.123	0.034	7.070
GCPM_2127-1	2.000	1.742	0.585	0.426	-0.372	2.000	1.676	0.347	0.403	0.140	3.000	1.944	0.618	0.486	-0.273	-0.179	-0.165	0.011	21.512
GCPM_672-1	10.000	7.504	0.545	0.867	0.371	10.000	6.178	0.930	0.838	-0.109	14.000	7.767	0.810	0.871	0.070	0.113	0.133	0.022	10.946
GCPM_3390-1	7.000	2.103	0.349	0.525	0.334	9.000	3.568	0.420	0.720	0.416	12.000	1.921	0.267	0.479	0.444	0.399	0.418	0.032	7.583
GCPM_1838	9.000	4.343	0.727	0.770	0.055	10.000	4.534	0.920	0.779	-0.180	11.000	5.487	0.873	0.818	-0.068	-0.065	-0.036	0.027	8.915
GCPM_4045-1	3.000	2.534	0.246	0.605	0.594	3.000	1.935	0.324	0.483	0.330	4.000	1.802	0.316	0.445	0.291	0.423	0.467	0.077	2.983
GCPM_4046-1	8.000	3.590	0.631	0.721	0.126	12.000	6.786	0.850	0.853	0.003	14.000	7.072	0.759	0.859	0.115	0.079	0.108	0.032	7.601
ORPM_162	2.000	1.813	0.179	0.448	0.602	2.000	1.992	0.031	0.498	0.937	2.000	1.959	0.274	0.489	0.440	0.663	0.667	0.012	20.523
GCPM_3586	4.000	2.562	0.385	0.610	0.369	3.000	1.819	0.041	0.450	0.909	4.000	2.884	0.167	0.653	0.745	0.654	0.668	0.041	5.857
GCPM_1676-1	6.000	3.582	0.833	0.721	-0.156	7.000	3.828	0.693	0.739	0.061	8.000	4.712	0.772	0.788	0.020	-0.023	-0.012	0.011	22.824
GCPM_2613-1	2.000	1.135	0.032	0.119	0.733	2.000	1.112	0.107	0.101	-0.056	2.000	1.027	0.026	0.026	-0.013	0.330	0.338	0.011	21.690
GCPM_3264-1	2.000	1.846	0.533	0.458	-0.164	2.000	1.813	0.226	0.449	0.495	2.000	1.933	0.407	0.483	0.157	0.160	0.163	0.004	70.527
GCPM_3937-1	3.000	1.578	0.303	0.366	0.173	3.000	1.719	0.520	0.418	-0.243	5.000	1.547	0.392	0.354	-0.109	-0.068	-0.062	0.005	45.346
PMGC_2522	5.000	3.656	0.603	0.727	0.170	5.000	3.859	0.813	0.741	-0.098	5.000	3.187	0.544	0.686	0.207	0.090	0.103	0.014	17.112
GCPM_1373-1	3.000	1.522	0.365	0.343	-0.064	2.000	1.805	0.671	0.446	-0.505	2.000	1.891	0.405	0.471	0.140	-0.144	-0.115	0.025	9.623
GCPM_2791-1	3.000	2.167	0.422	0.539	0.217	4.000	1.178	0.107	0.151	0.293	4.000	1.953	0.465	0.488	0.048	0.156	0.220	0.075	3.085
GCPM_4029-1	6.000	4.218	0.746	0.763	0.022	4.000	3.636	0.892	0.725	-0.230	7.000	4.105	0.797	0.756	-0.054	-0.085	-0.048	0.034	7.070
GCPM_2705-1	5.000	1.621	0.379	0.383	0.011	5.000	2.220	0.667	0.550	-0.213	5.000	2.711	0.620	0.631	0.017	-0.065	-0.026	0.037	6.599
PMGC_2558	4.000	2.089	0.656	0.521	-0.259	3.000	2.024	0.405	0.506	0.199	5.000	2.101	0.429	0.524	0.182	0.039	0.095	0.058	4.064
GCPM_2615-1	3.000	1.410	0.349	0.291	-0.201	3.000	2.122	0.439	0.529	0.169	3.000	1.977	0.486	0.494	0.017	0.030	0.109	0.081	2.839
ORPM_136	4.000	1.159	0.048	0.137	0.647	4.000	2.481	0.432	0.597	0.276	4.000	2.254	0.684	0.556	-0.229	0.098	0.197	0.110	2.024
Mean	4.885	2.911	0.474	0.548	0.138	5.000	2.908	0.485	0.567	0.152	6.000	3.288	0.472	0.591	0.170	0.150	0.182	0.038	12.818
SE	0.512	0.335	0.044	0.043	0.061	0.595	0.310	0.056	0.038	0.074	0.742	0.373	0.044	0.043	0.052	0.050	0.049	0.005	2.966

N_A = allele number; Ho = observed heterozygosity; H_E = expected heterozygosity; F_{IS} = fixation index in subpopulations; F_{IT} = fixation index in total population; F_{ST} = genetic differentiation of subpopulations; N_M = gene flow estimated from F_{ST}.

Population structure

The AMOVA indicated different levels of genetic variance among populations and among individuals within populations. Of the total genetic variance, 7% was attributed to population divergence, and the remaining 93% was explained by individual differences within populations (Table 4). The total F_{ST}, estimated at each locus for all individuals, ranged from 0.004 to 0.110 (mean 0.038), and the results were consistent with the history of high gene flow (mean 12.818).

Table 4. AMOVA analysis using 220 genotypes from three native *Populus talassica* populations.

Source	Degrees of freedom	Sum of squares	Mean square	Percentage variation
Among populations	2	253.655	126.828	7%
Within populations	217	4276.199	19.706	93%
Total	219	4529.855		100%

The results of pairwise population analysis using Nei's unbiased genetic distance showed that the largest F_{ST} was 0.100, between the Upper and Middle populations, and the smallest was 0.054, between the Middle and Lower populations (data not shown). Population structure analysis was conducted according to the known order of individuals, yielding the

 ©FUNPEC-RP www.funpecrp.com.br

optimal substructure result K = 2. Figure 2 illustrates the estimated subpopulations for the 220 individuals. Each individual is represented by a thin vertical line and classified based on its estimated membership probability (Q). Since STRUCTURE could not perform an analysis of K = 1 on populations with no difference, the optimal K = 2 is still unconvincing, especially considering the very low F_{ST} values and the 7% inter-population variance. No correlation was found between the genetic and geographic distances according to the Mantel test (data not shown).

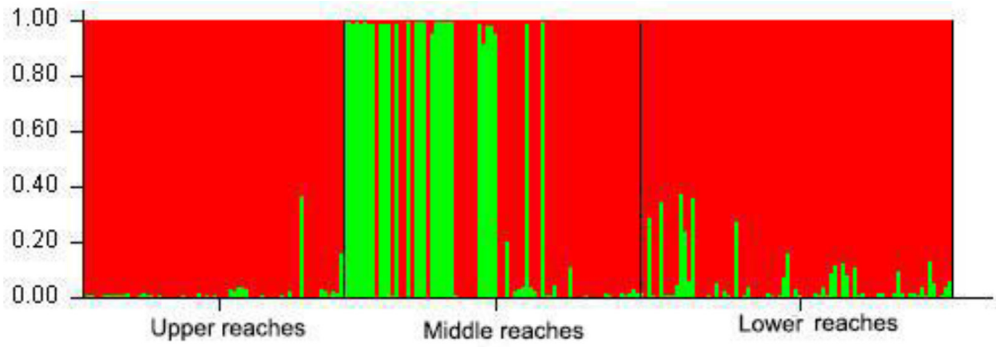

Figure 2. Population structure of *Populus talassica* estimated by STRUCTURE. In the figure, the sampling individuals were sorted as the three geographical sites.

DISCUSSION

Genetic diversity

To lay the foundation for association analysis and conservation management in *P. talassica*, we screened 26 highly polymorphic SSR primers and used them to evaluate genetic diversity and population structure. In the present study, we found a mean number of 6.629 observed alleles per locus within the 220 genotypes, belonging to three populations. This value was higher than the N_A (3.73) in the Tibetan poplar (*P. szechuanica* var. *tibetica*), which was determined using 24 SSR markers (Shen et al., 2014). Also, this value was higher than for some related *Populus* species identified using SSRs, such as *P. nigra*, *P. trichocarpa*, *P. tremuloides* (Slavov and Zhelev, 2010), and *P. tomentosa* (Du et al., 2012a,b). In our study, the mean H_O and H_E were 0.476 and 0.591, respectively. Similar values were also reported in other related species, such as *P. tremuloides* (H_O = 0.472, H_E = 0.67; Namroud et al., 2005) and *Populus tremula* (H_O = 0.474, H_E = 0.500; Lexer et al., 2005).

This relatively high N_A value may be related to the large sample size and co-dominant SSRs. As described in the Introduction, the 220 sampled genotypes represented a large proportion of the Kashi River region, which was the main area of distribution in the west of China. Thirteen primers had PIC values higher than 0.5, with a mean of 0.544 in 26 loci, both suggesting relatively high polymorphism. Another possible partial explanation for the considerable level of polymorphism is outcrossing, which is prevalent in *Populus* due to self-incompatibility.

Population structure

The STRUCTURE analysis revealed that the population should be divided into two subsets (K = 2). Individuals from upper and lower reaches were clustered into one group, and those from the middle clustered into another group. As Figure 2 shows, the Upper and Lower populations belonged to a single homogeneous population cluster. The Middle population showed mixed ancestry, suggesting a degree of genetic mixing. Since STRUCTURE could not perform analysis on K = 1, the result of K = 2 is unconvincing and debatable, especially considering the 7% among-population variance. The total F_{ST}, estimated at each locus for all individuals, was very low (mean 0.038), and the results of pairwise population analysis using Nei's unbiased genetic distance were also very low. Therefore, there was no significant differentiation among the three populations.

This may be mainly explained by strong gene flow and local geographic structure. Gene flow can lead to combining of gene pools, which can promote the reduction of genetic variation among groups (Slarkin, 1985). Gene flow also plays a considerable role in evolution by pollen dispersal, seed dispersal, and the establishment of the individual adults. This is especially true for high outcrossing and perennial species (Muona et al., 1990; Shen et al., 2014). In our study, a relatively strong gene flow was determined, with a mean Nm (the number of migrants successfully entering a population per generation) of 12.818. This value was much higher than for the related species *Populus alba* (mean $N_M = 3.1$; Lexer et al., 2005) and *P. szechuanica* (mean $N_M = 7.020$; Shen et al., 2014).

Another important factor contributing to the low level of differentiation among populations is local geographic structure, specifically, the river and the altitude. As mentioned in the Introduction, the three sampled populations were along the river basin, some in the river channel itself. Rivers cannot be ignored as an important factor in seed dispersal, and the river is an established seed dispersal mechanism of *P. talassica*. Another key factor that influences pollen and/or seed dispersal is altitude. The altitude gradient is a useful natural environment for investigating evolutionary responses to geophysical influences (Körner, 2007). For sampled populations that cover different altitudes, structural differentiation among populations could be due to restricted gene movement as a result of non-random mating and/or a geographic barrier (Byars et al., 2007; Pickup and Barrett, 2013). In the present study, the vertical distance was only 182 meters (m), with altitudes ranging from 1150 m (Upper) to 968 m (Lower). Therefore, we speculate that altitude had limited influence in gene flow in this study.

Conservation management

Future conservation management plans should aim to protect as many individuals as possible, as each individual transfers pollen and seeds. Rapid deforestation will greatly affect gene exchange between individuals and hinder pollen and/or seed dispersal. Consequently, this could lead to high levels of genetic differentiation between populations. Long-term geographic barriers also increase the probability of extinction, or they could lead to locally adapted populations. Long-term conservation should focus on protecting genetic diversity, not only individuals and/or populations.

CONCLUSIONS

To our knowledge, this is the first report of species-specific SSR markers and genetic

analysis in *P. talassica*. The genetic data revealed that the three populations along the Kashi River have relatively high diversity and a low level of genetic differentiation. A strong history of gene flow and the river itself play important roles in pollen and seed dispersal in this region. Therefore, conservation management should aim to protect as many individuals as possible.

ACKNOWLEDGMENTS

The authors are grateful to the anonymous reviewers for their constructive comments. Research partly supported by the National Natural Science Foundation of China "Molecular basis of the vegetative growth advantage in allotriploid poplar (#31530012), the Special Fund for Forest Scientific Research in the Public Welfare (#201404113), and the "111" Project (#B13007). We thank Chunbo Dong and Shiping Cheng for technical assistance with data analysis.

REFERENCES

Barrett SC and Kohn JR (1991). Genetic and evolutionary consequences of small population size in plants: implications for conservation. In: Genetics and conservation of rare plants (Falk DA and Holsinger KE, eds.). Oxford University Press, New York, 3-30.

Bo HX (2003). Study of the gene expression during the course of Na_2HCO_3 treatment in *P. talassica* Komm X *P. euphratica* Oliv using mRNA differential display. Master's thesis, Northeast Forestry University, Heilongjiang.

Byars SG, Papst W and Hoffmann AA (2007). Local adaptation and cogradient selection in the alpine plant, *Poa hiemata*, along a narrow altitudinal gradient. *Evolution* 61: 2925-2941. http://dx.doi.org/10.1111/j.1558-5646.2007.00248.x

Dong CB, Suo YJ and Kang XY (2014). Assessment of the genetic composition of triploid hybrid *Populus* using SSR markers with low recombination frequencies. *Can. J. For. Res.* 44: 692-699. http://dx.doi.org/10.1139/cjfr-2013-0360

Du Q, Wang B, Wei Z, Zhang D, et al. (2012b). Genetic diversity and population structure of Chinese White poplar (*Populus tomentosa*) revealed by SSR markers. *J. Hered.* 103: 853-862. http://dx.doi.org/10.1093/jhered/ess061

Du QZ, Zhang DQ and Li BL (2012a). Development of 15 novel microsatellite markers from cellulose synthase genes in *Populus tomentosa* (Salicaceae). *Am. J. Bot.* 99: e46-e48. http://dx.doi.org/10.3732/ajb.1100308

Earl DA and vonHoldt BM (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4: 359-361. http://dx.doi.org/10.1007/s12686-011-9548-7

Evanno G, Regnaut S and Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611-2620. http://dx.doi.org/10.1111/j.1365-294X.2005.02553.x

Hedrick PW and Kalinowski ST (2000). Inbreeding depression in conservation biology. *Annu. Rev. Ecol. Syst.* 31: 139-162. http://dx.doi.org/10.1146/annurev.ecolsys.31.1.139

Higgins K and Lynch M (2001). Metapopulation extinction caused by mutation accumulation. *Proc. Natl. Acad. Sci. USA* 98: 2928-2933. http://dx.doi.org/10.1073/pnas.031358898

Jakobsson M and Rosenberg NA (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23: 1801-1806. http://dx.doi.org/10.1093/bioinformatics/btm233

Keller LF and Waller DM (2002). Inbreeding effects in wild populations. *Trends Ecol. Evol.* 17: 230-241. http://dx.doi.org/10.1016/S0169-5347(02)02489-8

Körner C (2007). The use of 'altitude' in ecological research. *Trends Ecol. Evol. (Amst.)* 22: 569-574. http://dx.doi.org/10.1016/j.tree.2007.09.006

Hufford KM and Mazer SJ (2003). Plant ecotypes: genetic differentiation in the age of ecological restoration. *Trends Ecol. Evol.* 18: 147-155. http://dx.doi.org/10.1016/S0169-5347(03)00002-8

Lande R (1993). Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *Am. Nat.* 142: 911-927. http://dx.doi.org/10.1086/285580

Lande R (1994). Risk of population extinction from new deleterious mutations. *Evolution* 48: 1460-1469. http://dx.doi.org/10.2307/2410240

Lexer C, Fay MF, Joseph JA, Nica MS, et al. (2005). Barrier to gene flow between two ecologically divergent *Populus* species, *P. alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. *Mol. Ecol.* 14: 1045-1057. http://dx.doi.org/10.1111/j.1365-294X.2005.02469.x

Marshall TC, Slate J, Kruuk LE and Pemberton JM (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* 7: 639-655. http://dx.doi.org/10.1046/j.1365-294x.1998.00374.x

Muona O, Brown A, Clegg M, Kahler A, et al. (1990). Population genetics in forest tree improvement. *Plant Populat. Genet. Breed. Genet. Res.* 282-298.

Namroud MC, Park A, Tremblay F and Bergeron Y (2005). Clonal and spatial genetic structures of aspen (*Populus tremuloides* Michx.). *Mol. Ecol.* 14: 2969-2980. http://dx.doi.org/10.1111/j.1365-294X.2005.02653.x

Newman D and Pilson D (1997). Increased probability of extinction due to decreased genetic effective population size: experimental populations of Clarkia pulchella. *Evolution* 51: 354-362. http://dx.doi.org/10.2307/2411107

Newman D and Tallmon DA (2001). Experimental evidence for beneficial fitness effects of gene flow in recently isolated populations. *Conserv. Biol.* 15: 1054-1063. http://dx.doi.org/10.1046/j.1523-1739.2001.0150041054.x

Peakall R and Smouse PE (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6: 288-295. http://dx.doi.org/10.1111/j.1471-8286.2005.01155.x

Pickup M and Barrett SC (2013). The influence of demography and local mating environment on sex ratios in a wind-pollinated dioecious plant. *Ecol. Evol.* 3: 629-639. http://dx.doi.org/10.1002/ece3.465

Pritchard JK, Stephens M and Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

Rahman MH, Dayanandan S and Rajora OP (2000). Microsatellite DNA markers in *Populus tremuloides. Genome* 43: 293-297. http://dx.doi.org/10.1139/gen-43-2-293

Raymond M and Rousset F (1995). GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *J. Hered.* 86: 248-249.

Rousset F (1997). Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics* 145: 1219-1228.

Shen D, Bo W, Xu F and Wu R (2014). Genetic diversity and population structure of the Tibetan poplar (*Populus szechuanica* var. tibetica) along an altitude gradient. *BMC Genet.* 15 (Suppl 1): S1-S11. http://dx.doi.org/10.1186/1471-2156-15-S1-S11

Slarkin M (1985). Gene flow in natural populations. *Annu. Rev. Ecol. Syst.* 16: 393-430. http://dx.doi.org/10.1146/annurev.es.16.110185.002141

Slavov GT and Zhelev P (2010). Salient biological features, systematics, and genetic variation of *Populus*. In: Genetics and genomics of *Populus* (Jansson S, Bhalerao R, Groover AT, eds). NY: Springer. 15-38.

Tallmon DA, Luikart G and Waples RS (2004). The alluring simplicity and complex reality of genetic rescue. *Trends Ecol. Evol. (Amst.)* 19: 489-496. http://dx.doi.org/10.1016/j.tree.2004.07.003

Tóth G, Gáspári Z and Jurka J (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* 10: 967-981. http://dx.doi.org/10.1101/gr.10.7.967

Tuskan GA, Gunter LE, Yang ZK, Yin TM, et al. (2004). Characterization of microsatellites revealed by genomic sequencing of *Populus trichocarpa. Can. J. For. Res.* 34: 85-93. http://dx.doi.org/10.1139/x03-283

Van Oosterhout CV, Hutchinson WF, Wills DPM and Shipley P (2004). Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* 4: 535-538. http://dx.doi.org/10.1111/j.1471-8286.2004.00684.x

Volis S and Blecher M (2010). Quasi *in situ*: a bridge between *ex situ* and *in situ* conservation of plants. *Biodivers. Conserv.* 19: 2441-2454. http://dx.doi.org/10.1007/s10531-010-9849-2

Yin TM, Zhang XY, Gunter LE, Li SX, et al. (2009). Microsatellite primer resource for *Populus* developed from the mapped sequence scaffolds of the Nisqually-1 genome. *New Phytol.* 181: 498-503. http://dx.doi.org/10.1111/j.1469-8137.2008.02663.x

Zhang DY, Cui PY, Tao XD and Guo JZ (2010). Hybrid breeding on *Populus talassica* X *P. euphratic* and approach of test and study on salt resistance. *Xinjiang Agric. Sci* 47: 246-250.