Development of Microsatellite Markers Using Next-Generation Sequencing and Genetic Characterization in Three Natural Monument Populations of *Koelreuteria paniculata* (Goldenrain Tree)

Jei-Wan Lee*, Sang-Chul Kim, Sookyung Shin, Ji-Young Ahn, and Min-Woo Lee

Department of Forest Bioresources, National Institute of Forest Science, Suwon 16631, Korea

*Corresponding author: leejeiwan@korea.kr

Abstract

This study was conducted to develop microsatellite markers in *Koelreuteria paniculata* using next-generation sequencing. A total of 71,114,562 reads of 20x coverage for the *K. paniculata* genome were generated and assembled to 141,924 contigs of a minimum of 500 bp long. One hundred seventy-nine of the 79,633 contigs containing microsatellite regions were used for primer design. Fourteen primer sets were selected as polymorphic markers by applying them to three *K. paniculata* populations, designated as natural monuments in Korea. Somewhat low levels of genetic diversity were observed. The grand means of the number of alleles, observed heterozygosity, and expected heterozygosity were 2.7, 0.493, and 0.407, respectively. Analysis of molecular variance indicated a high level of genetic differentiation among populations. Distinct patterns of three populations were identified in the principal coordinate analysis and Bayesian structure analysis. The probability of identity of these markers was estimated to be quite low, suggesting that these markers have the robust potential power to distinguish genetically different individuals. *K. paniculata* has a high economic value as an ornamental tree, and a honey tree. The novel microsatellite markers developed in this study will be useful for future breeding programs and genetic studies aimed at developing conservation plans.

Additional key words: genetic diversity, genetic structure, individual identification, ornamental tree, simple sequence repeat

Introduction

Koelreuteria paniculata Laxm. (Sapindaceae) is a deciduous woody plant, reaching up to 13 m in height and 50 cm in diameter (Meyer, 1976). This species is regionally endemic and native to parts of eastern Asia including Korea, China, and Japan (Wang et al., 2013). It has high value as a horticultural genetic resource because of the aesthetic appeal of its showy yellow flowers and seed pods. It has been widely used as an ornamental and landscape tree at parks, roadsides, temples, and so on, and has been introduced as the 'Goldenrain tree' in America and Europe, and is popularly cultivated there for

landscape purposes. In addition, its fruits and leaves have been used as medicinal resources and dyes, respectively (Meyer, 1976). In South Korea, *K. paniculata* trees are scattered sporadically with only a few remnant populations, mainly located near the seaside (Lee, 1958; Lee et al., 1997). Among them, three populations have been designated as natural monuments due to their biological and cultural significance. These remnant populations may have important evolutionary roles as reservoirs of genetic diversity and reproductive fitness, and for maintaining the adaptive potential of trees under pressure from environmental change (Reisch et al., 2007). In that respect, those natural monument populations could play an important role as a genetic source for conserving genetic diversity and developing new economic varieties of this species. For the conservation of the monument populations, local authority-led management projects, such as tree surgery and removal of understory vegetation, have been carried out. Meanwhile, it is also necessary to prepare conservation strategies that consider the genetic characteristics of these populations. Genetic studies of *K. paniculata* in Korea have been performed by isozyme analysis (Lee et al., 1997). However, studies using DNA markers, based on sequence information such as microsatellites, have not been conducted.

A microsatellite (or simple sequence repeats, SSR) is a highly variable genetic region consisting of simple sequence repeats of 1 to 6 nucleotides (Selkoe and Toonen, 2006). Microsatellites are widely distributed in genomes and have been used as a genetic marker for studying genetic diversity, breeding, forensic science, and so on (Pourkhaloee et al., 2018; Kim et al., 2019). Ten polymorphic expressed sequence tag (EST)-SSR markers were previously developed using the transcriptomic approach in this species (Yang et al., 2017). However, the results showed slightly low polymorphism. In addition, since EST-SSRs are only derived from transcribed regions and likely to be conserved, they are less randomly dispersed throughout the genome and less polymorphic than genomic microsatellites (Li et al., 2004; Pashley et al., 2006). Employing a large number of markers can allow for genome-wide sampling and increase the statistical power for genetic estimation (Selkoe and Toonen, 2006). Thus, more DNA markers need to be developed for the genetic characterization of this species.

In this study, we aimed to develop polymorphic genomic microsatellite markers for *K. paniculata* using next-generation sequencing, and to characterize them using the three natural monument populations of *K. paniculata* in Korea.

Materials and Methods

Next-Generation Sequencing and Assembly

To construct the DNA library for next-generation sequencing, DNA was extracted from an individual *K. paniculata* tree that is planted in the National Institute of Forest Science (NIFoS), Suwon, South Korea (voucher specimen No. WFRI 67978; deposited in the herbarium at the Warm-temperate and Subtropical Forest Research Center of NIFoS, JeJu). Genomic DNA (100 ng·μL^{-1}) was extracted using Plant SV mini kit (GeneAll, Seoul, Korea) and digested using Fragmentase (New England Biolabs, Beverly, MA, USA). The fragments were repaired to blunt ends using a SPARK DNA sample Prep Kit for Ion Torrent (Enzymatics, Beverly, MA, USA). The P1 adapter (5'-CCACTACGCCTCCGCT-TTCCTCTCTATGGGCAGTCGGTGAT-3', 5'-ATCACCGACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTG-GTT-3') and A adapter (5'-CTGAGTCGGAGACACGCAGGGATGAGATGGTT-3', 5'-CCATCTCATCCCTGCGTG-TCTCCGACTCAG-3') were ligated to the blunt-ended DNA using SPARK DNA sample Prep Kit (Enzymatics). DNA

fragments of 230-270 bp were selected by purification using XSEP MagBead (Celemics, Seoul, Korea). The purified DNA was used as templates in PCR amplification using the primer set (5'-CCACTACGCCTCCGCTTTCCTCTCTATG-3', 5'-CCATCTCATCCC TGCGTGTC-3'). The PCR product was applied to next-generation sequencing with an Ion Proton system (Thermo Fisher Scientific, Waltham, MA, USA). The sequenced reads were assembled to contigs using SOAP-denovo2 ver. 2.04 (Luo et al., 2012).

Microsatellite Detection, Primer Design, and PCR Amplification

To identify the microsatellites, the contigs larger than 500 bp were selected and analyzed by MISA (Beier et al., 2017). The length of the repeated motif for detecting microsatellites was set from di- to penta-nucleotides. The minimum numbers of repetitions were 6 for di-nucleotides, and 5 for the others. Next, primer sets flanking the microsatellites were designed with Primer3 (Koressaar and Remm, 2007).

To select stable and polymorphic candidate markers, the primer sets were screened using 12 individuals (four per population) collected from the three natural monument populations located in Anmyeondo (AM), Pohang (PH), and Wando (WD) in South Korea (Table 1 and Fig. 1). Then, selected polymorphic markers were applied to all individuals sampled from the three populations. The PCR amplification was performed in 15 µL reaction mixtures containing the

Table 1. Location information for *K. paniculata* samples used in this study

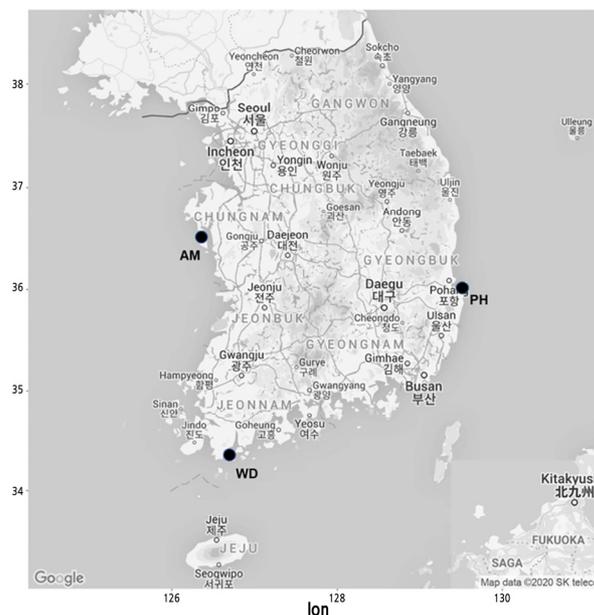| Population	Locality	Natural monument No.	Geographic coordinates	Number of individuals
Anmyeondo (AM)	Anmyeon-eup, Taean-gun, Chungcheongnam-do	138	36° 30' N, 126° 20' E	30
Pohang (PH)	Balsan-ri, Pohang-si, Gyeongsangbuk-do	371	36° 02' N, 129° 30' E	30
Wando (WD)	Wando-eup, Wando-gun, Jeollanam-do	428	34° 21' N, 126° 38' E	48

Fig. 1. Locations of the three natural monument populations of *K. paniculata* analyzed in this study. Locations are indicated by black solid circles.

following: 20 ng of template DNA, 1x reaction buffer, 2.5 mM MgCl$_2$, 0.2 mM dNTPs, 0.04 μM M13(-19) sequencing primer labelled with FAM, 0.2 μM primer mix, and 0.5 U of *Taq* DNA polymerase (BIOFACT, Daejeon, Korea). The PCR conditions were as follows: denaturing at 94°C for 5 min; 10 cycles at 94°C for 45 s, 60°C for 45 s, and 72°C for 45 s; 30 cycles at 94°C for 45 s, 52°C for 45 s, and 72°C for 45 s; and final extension at 72°C for 5 min. The PCR products were separated on an ABI 3730 *xl* Genetic Analyzer (Applied Biosystems), and genotypes were determined using Gene Mapper v5.0 (Applied Biosystems).

Polymorphism and Genetic Characterization

Genetic parameters such as the number of alleles (N_A), observed heterozygosity (H_O), expected heterozygosity (H_E), inbreeding coefficient (F_{IS}), pairwise genetic differentiation index (F_{ST}), and accumulated probability of identity (PI) for polymorphic markers were calculated using GenAlEx 6.41 (Peakall and Smouse, 2006). Analysis of molecular variance (AMOVA) for estimating the components of variance of the populations, and principal coordinate analysis (PCoA) for detecting the pattern of genetic differentiation of the populations were conducted using the GenAlEx. The Hardy-Weinberg (HW) equilibrium was tested by the Fisher's exact test based on a Markov chain (Forecasted chain length: 1,000,000 and Dememorization step: 100,000) for all loci in each population using Arlequin ver. 3.5 (Excoffier and Lischer, 2010). The linkage disequilibrium (LD) was tested using FSTAT ver. 2.9.3.2 (Goudet, 2002). The significance levels were adjusted using the Bonferroni correction. Null allele frequencies were estimated using the Micro-Checker 2.2.3 (van Oosterhout et al., 2004). The polymorphic information content (PIC) was calculated using PIC calculator (Jan, 2002), according to the equation:

$$PIC = 1 - \sum_{i=1}^{n} p_i^2 - 2\left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} p_i^2 p_j^2\right]$$

where p_i is the frequency of the ith allele, and n is the number of alleles (Botstein et al., 1980). The population structure was inferred using STRUCTURE ver. 2.3.4 based on an admixture model with correlated allele frequencies. The K value was set from 1 to 6. For each K value, 10 independent runs were conducted with a burn-in period of 100,000 for each run, and a Markov chain Monte Carlo of 100,000 iterations. The optimum K value was determined by the method of Evanno et al. (2005) using STRUCTURE HARVESTER ver. 0.6.94 (Earl and Vonholdt, 2012).

Results and Discussion

Sequencing, Contig Assembly, and Microsatellite Detection

The next-generation sequencing analysis obtained 71,114,562 reads with a total length of 10.3 Gbp (average 144 bp), accounting for about 20x genome coverage (Ohri et al., 2004), which were assembled into 141,924 contigs of a minimum of 500 bp long with an average length of 784 bp (Table 2). The average length was quite similar to that of *Haloxylon ammodendron* analyzed with the Ion Proton, as in this study (Batkhuu et al., 2019). Among them, 141,441 contigs (99.7%) were identified as derived from the nuclear genome by BlastN analysis, and included 15,042 microsatellites. There were 5,699 and 4,175 di- and tri-nucleotide motifs, respectively, with the most abundant accounting for 66% (Fig. 2).

Table 2. Results of next-generation sequencing and assembly

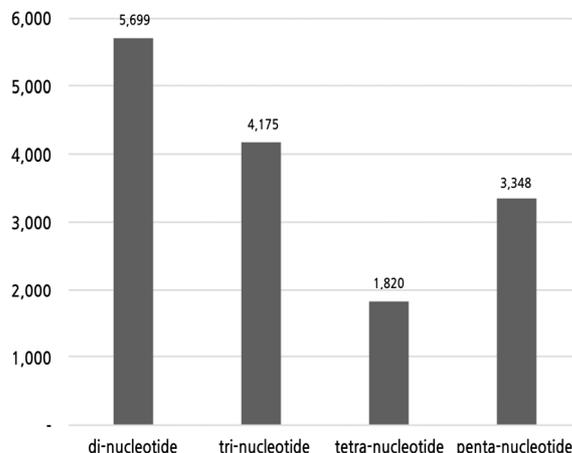| Number of reads	Total length	Number of contigs	Length of contigs (bp)				
			Total length	Minimum	Maximum	Average	N50
71,114,562	10.3 Gbp	141,924	111,368,532	500	6,057	784	779

Fig. 2. Distribution of microsatellite repeat motifs.

Penta- and tetra-nucleotide microsatellites were 3,348 and 1,820, respectively, accounting for 34%. These results were consistent with most other organisms, excluding mono-nucleotide motifs which were not considered for analysis in this study (Zhang et al., 2012; Baek et al., 2016; Batkhuu et al., 2019). All of the 163 primer sets were designed from 179 contigs containing di- and tri-nucleotide microsatellites.

Characterization of Polymorphic Markers and Genetic Diversity

Fourteen primer sets were selected as polymorphic markers based on the genotyping data of the three *K. paniculata* populations (Table 3). The sequences including polymorphic microsatellite regions were deposited in the NCBI GenBank. The number of alleles at each locus for all populations ranged from one to seven (Table 4). The null allele frequency (*NAF*) of polymorphic loci were estimated to be negligible ($p < 0.05$) or moderate ($0.05 < NAF < 0.20$; Chapuis and Estoup, 2007). No significant LD value ($p < 0.01$) was detected in all marker combinations. The average number of alleles ranged from 2.4 in WD to 2.9 in AM. These results were similar to those analyzed in the previous study reported by Yang et al. (2017), in which ten polymorphic microsatellite markers were developed using the transcriptome data of *K. paniculata*. Genomic microsatellite markers were generally expected to have higher polymorphism compared to EST-derived microsatellite markers designed in highly conserved regions (Pashley et al, 2006). However, in this study, there was no difference in polymorphism depending on the sequence source for marker development. This low level of allelic diversity in *K. paniculata* was similar to that of *Calophyllum inophyllum* (Setsuko et al., 2012). These two species are distributed along the coast, and their seeds are dispersed through the ocean currents to areas relatively far apart (Lee, 1958). Setsuko et al. (2012) explained that the low allelic diversity of *C. inophyllum* might be due to founder effects and genetic drift in the process of the introduction of seeds by the ocean current (Eckert et al., 1996). We also thought that the low allelic diversity of *K. paniculata* might be due to the founder effect, and genetic drift by such a seed dispersal mechanism.

The average observed heterozygosity ranged from 0.406 in AM to 0.581 in PH, and the average expected heterozygosity ranged from 0.372 in AM to 0.455 in PH (Table 4). These results tended to be slightly lower compared to other tree species with a similar life history (long-lived perennial: $H_E = 0.680$, regional: $H_E = 0.650$, outcrossing: $H_E = 0.650$) (Nybom,

Table 3. Characteristics of 14 microsatellite markers developed in *K. paniculata*

Locus	Primer sequence (5'-3')		Repeat motif	Size range (bp)	GenBank accession No.
MGJ_012	F: GTGGACCACAAAACATGTGC	R: TTATTGTGGCAAAATGGAAGG	(CA)14	195 – 205	MK937231
MGJ_016	F: GAATGTGGCATAACCTCTTGG	R: GGATTGTGTGTAAGTGTAAGTTGAG	(CT)14	242 – 258	MK937232
MGJ_019	F: CCAAGCCATAGCCTCTTCAG	R: GCATTCACGCTGTTTTAACG	(GA)11	241 – 255	MK937233
MGJ_021	F: AATGAGGAGGGAGGAAGAGC	R: AACGGCGTATGATGGAACAC	(CT)10	185 – 191	MK937234
MGJ_041	F: CCTGCAGGTAGGCTAAGGTG	R: GAGTTTGTCTTTTACAAGGAGTGC	(TC)10	206 – 214	MK937235
MGJ_045	F: TCCTCGAGTCGAAAGGTCAC	R: ATCGTCGGTCCAACTATTGC	(CT)10	216 – 224	MK937236
MGJ_067	F: ACCGGGTCGAACAGATAATG	R: GAATGGAGAAACCCTGATCG	(AAG)7	233 – 236	MK937237
MGJ_071	F: CCATTCCATTGTGCTTTTCAC	R: CAACTTGGCCAACCTCTTG	(TGA)8	244 – 265	MK937238
MO_105	F: GTGTTAGAAATTAGGGTTAGGG	R: CGACTAGTTCCGGGTTCGAG	(AAG)4	147 – 150	MK937241
MO_216	F: AAATTGGAGGAGCCATTCAG	R: TTGTTACCTGTGGCATGTCG	(AG)7	321 – 327	MK937242
MO_217	F: AACGTGGAATTGGTCTTCTTTC	R: AATCTGAATTTGACTCGTCTGC	(AT)6	378 – 392	MK937243
MO_230	F: GGTCAACTATCATCTCTAAAGG	R: TATCCCACCTTAATCCCAAC	(AT)9	203 – 237	MK937244
MO_238	F: TGTCAATGGAAAACTCCAAGC	R: TTCCTCATCACCGGAGGTC	(AT)7	267 – 277	MK937245
MO_261	F: AAAGCCCCAAATCTCACCTC	R: TGCTATGACAGGGTTGAAAGG	(CTT)4	310 – 337	MK937246

Table 4. Genetic parameters of 17 loci in the three natural monument populations of *K. paniculata*

Locus	AM population ($n = 30$)					PH population ($n = 30$)					WD population ($n = 48$)					PIC ($n = 108$)
	N_A	H_O	H_E	F_{IS}[z]	NAF	N_A	H_O	H_E	F_{IS}[z]	NAF	N_A	H_O	H_E	F_{IS}[z]	NAF	
MGJ_012	4	0.667	0.652	-0.023^*	-0.041	4	0.833	0.727	-0.146^{ns}	-0.081	3	0.548	0.601	0.087^{ns}	0.058	0.647
MGJ_016	5	0.556	0.603	0.078^*	0.060	5	0.538	0.595	0.096^{ns}	0.050	2	0.227	0.375	0.394^*	0.167	0.621
MGJ_019	3	0.400	0.335	-0.194^{ns}	-0.215	5	0.692	0.676	-0.024^{***}	-0.027	3	0.545	0.497	-0.098^{ns}	-0.049	0.616
MGJ_021	3	0.448	0.432	-0.037^{ns}	0.006	2	0.429	0.375	-0.143^{ns}	-0.079	3	0.395	0.383	-0.032^{ns}	-0.020	0.376
MGJ_041	3	0.367	0.395	0.072^{ns}	0.020	3	0.786	0.554	-0.419^{***}	-0.306	3	0.444	0.432	-0.028^{ns}	-0.018	0.394
MGJ_045	1	N/A	N/A	N/A	N/A	2	0.633	0.499	-0.268^{ns}	-0.145	2	0.489	0.470	-0.040^{ns}	-0.020	0.368
MGJ_067	1	N/A	N/A	N/A	N/A	2	0.400	0.464	0.139^{ns}	0.067	2	0.174	0.159	-0.095^{ns}	-0.091	0.215
MGJ_071	2	0.138	0.128	-0.074^{ns}	-0.072	2	0.345	0.285	-0.208^{ns}	-0.191	3	0.576	0.506	-0.137^{ns}	-0.080	0.391
MO_105	2	0.300	0.255	-0.176^{ns}	-0.163	2	0.846	0.488	-0.733^{***}	-0.608	2	0.976	0.500	-0.953^{***}	-0.846	0.372
MO_216	2	0.296	0.346	0.143^{ns}	0.066	2	0.815	0.494	-0.650^{***}	-0.423	1	N/A	N/A	N/A	N/A	0.252
MO_217	2	0.321	0.270	-0.191^{ns}	-0.176	3	0.107	0.103	-0.043^{ns}	-0.054	3	0.935	0.607	-0.541^{***}	-0.541	0.391
MO_230	7	0.464	0.702	0.339^{***}	0.182	1	N/A	N/A	N/A	N/A	3	0.045	0.088	0.482^{***}	0.167	0.448
MO_238	3	0.769	0.589	-0.307^{***}	-0.162	3	0.846	0.607	-0.393^{***}	-0.234	2	1.000	0.500	-1.000^{***}	-1.000	0.573
MO_261	2	0.962	0.499	-0.926^{***}	-0.804	2	0.862	0.500	-0.724^{***}	-0.475	2	0.534	0.387	-0.355^*	-0.310	0.365
Mean	2.9	0.406	0.372	-0.108	-0.093	2.7	0.581	0.455	-0.271	-0.179	2.4	0.492	0.393	-0.178	-0.184	0.431

n = number of individuals; N_A = number of alleles; H_O = observed heterozygosity; H_E = expected heterozygosity; F_{IS} = inbreeding coefficient; NAF = null allele frequency; P_{IC} = polymorphic information content; N/A = not available.

[z] Significant deviations from Hardy-Weinberg equilibrium: $*p < 0.05$, $**p < 0.01$; $***p < 0.001$, ns = not significant.

2004). The polymorphic information content (*PIC*) varied between loci from 0.215 (MGJ_067) to 0.647 (MGJ_012), with an average of 0.431 over all 14 loci (Table 4). Four loci were highly informative (*PIC* > 0.5), and nine loci were moderately informative (0.25 < *PIC* < 0.5) (Bostein et al., 1980). The accumulated probabilities of identity (*PI*) of the 14 markers were calculated to be 4.1×10^{-6}, 5.6×10^{-7}, and 7.5×10^{-6} in AM, PH, and WD populations, respectively. These results mean that the probability that two independent samples will have the same identical genotype is quite low, and the set of these loci has a robust potential power to distinguish genetically different individuals in the population (Peakall and Smouse, 2006).

Genetic Differentiation and Structure of Three *K. paniculata* Populations

AMOVA revealed that 69% of the total variation originated from the difference among individuals within populations, and that 31% was due to differences among populations (Table 5), which was similar to the results from *Hibiscus tiliaceus* (31.3%) (Takayama et al., 2008), and much higher than that of *C. inophyllum* (14.5%) (Hanaoka et al., 2014). In the previous study on the *K. paniculata* population in Korea based on isozyme analysis, a high level of genetic differentiation was also found compared to tree species with similar distribution (Lee et al., 1997). PCoA at the population level showed a clearly distinct pattern. The first two principal coordinates explained 65.22% and 34.78%, respectively (Fig. 3). Pairwise F_{ST} values ranged from 0.101 to 0.168 (Table 6). Analysis with the Evanno et al. (2005) method based on STRUCTURE outputs showed that the value of $\triangle K$ was the largest at $K = 3$, as the most likely number of clusters (Fig. 4). The high genetic differentiation and distinct pattern among three *K. paniculata* populations indicated that gene flow might be quite

Table 5. Distribution of genetic variations of three *K. paniculata* populations from analysis of molecular variance (AMOVA)

Source of variation	Degree of freedom	Sum of squares	Variance components	Percentage of variance (%)
Among populations	2	257.779	3.460	31.0
Within populations	105	819.388	7.804	69.0
Total	107	1077.167	11.263	100.0

Table 6. Pairwise genetic differentiation index values among the three populations

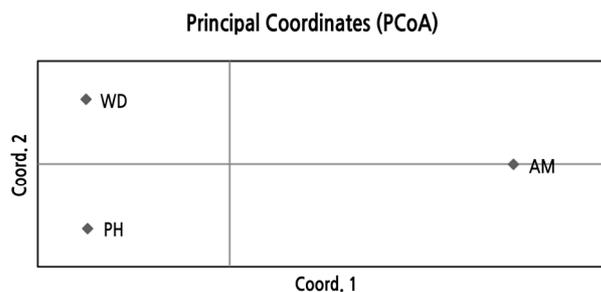| Populations	AM	PH	WD
AM	0.000	-	-
PH	0.166	0.000	-
WD	0.168	0.101	0.000

Principal Coordinates (PCoA)

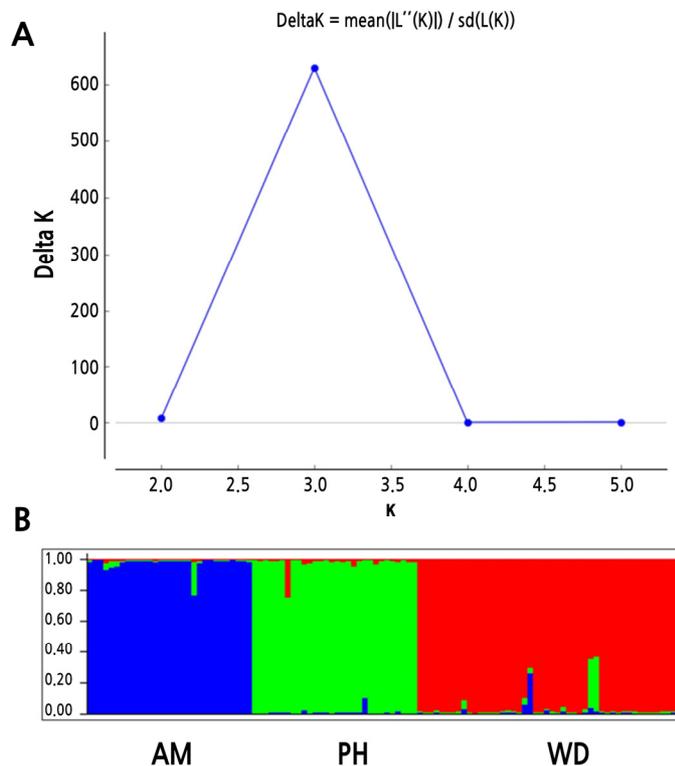Fig. 3. Principal components analysis (PCoA) for the three *K. paniculata* populations.

Fig. 4. Results of STRUCTURE analysis. (A) Estimation of ΔK based on LnP(D) for determining the most likely number of clusters; (B) Bar plots of the STRUCTURE analysis with $K = 3$.

rare among these populations. That might be due to its characteristic seed dispersal by gravity and drifting on the water, which is not enough for gene flow among populations discontinuously distributed over relatively long distances (Nybom, 2004). In fact, these populations are fragmented along the coast, and are at least about 230 km apart from each other when measured in a straight line (Fig. 1).

In conclusion, this study provided information regarding novel microsatellite markers and their genetic characteristics analyzed in the three natural monument populations of *K. paniculata* in Korea. Although the analyzed populations are designated as natural monuments and managed by legal protection, their genetic diversity falls short of expectations. This means that appropriate conservation strategies are needed to maintain the genetic diversity of these populations. Considering low levels of allelic diversity and high levels of genetic differentiation, *ex situ* conservation to collect the genotypes including all alleles distributed in each population are recommended. On the other hand, in order to prevent the reduction of genetic diversity due to genetic drift, introducing new genotypes from other populations also needs to be considered. In addition to the populations analyzed here, several populations are sporadically distributed inland. More detailed studies including the inland populations are expected to lead to appropriate measures to conserve genetic resources of *K. paniculata* in Korea. In this respect, microsatellite markers developed in this study will be useful for further conservational genetic studies. In addition, *K. paniculata* is one of the economically important species as an ornamental tree and a honey tree. Recently, many efforts are being made to select more suitable individuals according to their intended use. In this process, these markers will be effectively applicable for identifying selected breeding lines, individuals, and cultivars.

Literature Cited

Baek SH, Lee JW, Hong KN, Lee SW, Ahn JY, Lee MW (2016) Identification and characterization of polymorphic microsatellite loci using next generation sequencing in *Quercus variabilis*. J Korean For Soc 105:186-192. doi:10.14578/jkfs.2016.105.2.186

Batkhuu NO, Kim SC, Lee JW, Hong KN (2019) Development of 15 novel microsatellite markers for a *Haloxylon ammodendron* (Amaranthaceae) using next-generation sequencing. J For Res 24:382-385. doi:10.1080/13416979.2019.1675253

Beier S, Thiol T, Münch T, Scholz U, Mascher M (2017) MISA-wad: a web server for microsatellite prediction. Bioinformatics 33:2583-2585. doi:10.1093/bioinformatics/btx198

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314-331

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. Mol Biol Evol 24:621-631. doi:10.1093/molbev/msl191

Earl DA, Vonholdt BM (2012) Structure Harvester: a website and program for visualizing structure output and implementing the Evanno method. Conserv Genet Resour 4:359-361. doi:10.1007/s12686-011-9548-7

Eckert CG, Manicacci D, Barrett SC (1996) Genetic drift and founder effect in native versus introduced populations of an invading plant, *Lythrum salicaria* (Lythraceae). Evolution 50:1512-1519. doi:10.1111/j.1558-5646.1996.tb03924.x

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

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564-567. doi:10.1111/j.1755-0998.2010.02847.x

Goudet J (2002) FSTAT, a program to estimate and test gene diversities and fixation indices. http://www2.unil.ch/popgen/softwares/fatat.htm

Hanaoka S, Chien CT, Chen SY, Watanabe A, Setsuko S, Kato K (2014) Genetic structures of *Calophyllum inophyllum* L., a tree employing sea-drift seed dispersal in the northern extreme of its distribution. Ann For Sci 71:575-584. doi:10.1007/s13595-014-0365-5

Jan SJK (2002) PIC calculator. https://www.liverpool.ac.uk/~kempsj/pic.html

Kim HJ, Park SH, Kim JH, Yim B, Mun JH, Kim HB, Hur YY, Yu HJ (2019) An efficient strategy for developing genotype identification markers based on simple sequence repeat in grapevine. Hortic Environ Biotechnol 60:363-372. doi:10.1007/s13580-019-00123-x

Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. Bioinformatics 23:1289-1291. doi:10.1093/bioinformatics/btm091

Lee SW, Kim SC, Kim WW, Han SD, Yim KB (1997) Characteristics of leaf morphology, vegetation and genetic variation in the endemic populations of a rare tree species, *Koelreuteria paniculata* Laxm. J Korean For Soc 86:167-176

Lee YN (1958) Transportation of *Koelreuteria paniculata* by sea current. J Plant Biol 1:11-20

Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structures, function, and evolution. Mol Biol Evol 21:991-1007. doi:10.1093/molbev/msh073

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, et al. (2012) SOAPdenoveo2: an empirically improved memory-efficient short-read *de novo* assembler. Gigascience 1:18. doi:10.1186/2047-217X-1-18

Meyer F (1976) A revision of the genus *Koelreuteria* (Sapindaceae). J Arnold Arbor 57:129-166. Retrieved from http://www.jstor.org/stable/43782511

Nybom H (2004) Comparison of different nuclear DNA markers for estimation intraspecific genetic diversity in plants. Mol Ecol 13:1143-1155. doi:10.1111/j.1365-294X.2004.02141.x

Ohri D, Bhargava A, Chatterjee A (2004) Nuclear DNA amounts in 112 species of tropical hardwoods - New estimates. Plant Biol 6:555-561. doi:10.1055/s-2004-821235

Pashley CH, Ellis JR, McCauley DE, Burke JM (2006) EST databases as a source for molecular markers: Lessons from *Helianthus*. J Hered 97:381-388. doi:10.1093/jhered/esl013

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

Pourkhaloee A, Khosh-Khui M, Arens P, Salehi H, Razi H, Niazi A, Afsharifar A, van Tuyl J (2018) Molecular analysis of genetic diversity, population structure, and phylogeny of wild and cultivated tulips (*Tulipa* L.) by genic microsatellites. Hortic Environ Biotechnol 59:875-888. doi:10.1007/s13580-018-0055-6

Reisch C, Mayer F, Rüther C, Nelle O (2007) Forest history affects genetic diversity - molecular variation of *Dryopteris dilatata* (Dryopteridaceae) in ancient and recent forest. Nord J Bot 25:366-371. doi:10.1111/j.0107-055X.2008.00188.x

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol Lett 9:615-629. doi:10.1111/j.1461-0248.2006.00889.x

Setsuko S, Uchiyama K, Sugai K, Hanaoka S, Yoshimaru H (2012) Microsatellite markers derived from *Calophyllum inophyllum* (Clusiaceae) expressed sequence tags. Am J Bot 99:e28-e32. doi:10.3732/ajb.1100299

Takayama K, Tateishi Y, Murata JIN, Kajita T (2008) Gene flow and population subdivision in a pantropical plant with sea-drifted seeds *Hibiscus tiliaceus* and its allied species: Evidence from microsatellite analysis. Mol Ecol 17:2730-2742. doi:10.1111/j.1365-294X.2008.03799.x

van Oosterhout C, Hutchinson WF, Wills DMP, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535-538. doi:10.1111/j.1471-8286.2004.00684.x

Wang Q, Manchester SR, Gregor HJ, Shen S, Li ZY (2013) Fruit of *Koelreuteria* (Sapindaceae) from the Cenozoic throughout the northern hemisphere: their ecological, evolutionary, and biogeographic implications. Am J Bot 100:422-449. doi:10.3732/ajb.1200415

Yang X, Gao K, Chen Z, Yang X, Rao P, Zhao T, An X (2017) Development and application of EST-SSR markers in *Koelreuteria paniculata* Laxm. using a transcriptomic approach. Biotechnology 16:45-56. doi:10.3923/biotech.2017.45.56

Zhang Q, Ma B, Li H, Chang Y, Han Y, Li J, Wei G, Zhao S, Khan M, et al. (2012) Identification, characterization, and utilization of genome-wide simple sequence repeats to identify a QTL for acidity in apple. BMC Genom 13:537-10.1186/1471-/2164-13-537. doi:10.1186/1471-2164-13-537