JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2011, Vol. 56, No. 5
doi: 10.1111/j.1556-4029.2011.01810.x
Available online at: onlinelibrary.wiley.com

PAPER

CRIMINALISTICS

Eun Jung Lee,[1] *Ph.D.; Gang Nam Jin,*[1] *M.D.; Kyung Lyong Lee,*[1] *Ph.D.; Myun Soo Han,*[2] *Ph.D.; Yang Han Lee,*[2] *Ph.D.; and Moon Sik Yang,*[3] *Ph.D.*

Exploiting Expressed Sequence Tag Databases for the Development and Characterization of Gene-Derived Simple Sequence Repeat Markers in the Opium Poppy (*Papaver somniferum* L.) for Forensic Applications

ABSTRACT: Simple sequence repeat (SSR) markers in the opium poppy (*Papaver somniferum* L.) were identified from an expressed sequence tag (EST) database comprised of 20,340 sequences. In total, 2780 SSR-containing sequences were identified. The most frequent microsatellite had an AT/TA motif (37%). Twenty-two opium poppy EST-SSR markers were presently developed and polymorphisms of six markers (psom 2, 4, 12, 13, 17, and 22) were utilized in 135 individuals under narcotic control investigation. An average of three alleles per locus (range: 2–5 alleles) with a mean heterozygosity of 0.167 was detected. Six loci identified 29 unique profiles in 135 individuals. The EST-SSR markers exhibited small degrees of genetic differentiation (fixation index = 0.727, $p < 0.001$). Other variable markers will be needed to facilitate the forensic identification of the opium poppy for future cases. To determine the potential for cross-species amplification, six markers were tested in five *Papaver* genera species and two *Eschscholzia* genera. The psom 4 and psom 17 primer pair was transferable. This is the first study to report SSR markers of the opium poppy.

KEYWORDS: forensic science, *Papaver somniferum* L., expressed sequence tag, simple sequence repeats, microsatellites, diversity analysis, narcotic seizure

The opium poppy, *Papaver somniferum* L., is thought to have originated from the Mediterranean region and is one of the earliest medicinal plants known to mankind. Opium and many refined opiates, including morphine, thebaine, codeine, papaverine, and noscapine, are extracted from poppies. Drug ingredients such as morphine and codeine are tested using various chemical methods from drug users and poppy plants (1–3). *P. somniferum* L. is a species of plant with diverse characteristics of color, number and shape of petals, number of pods, and production of morphine (4). The opium poppy is widely grown in ornamental gardens throughout Europe, North America, South America, and Asia, but the possession and cultivation of this plant is outlawed in South Korea and in other countries because of its addictively intoxicating properties.

Microsatellites, or simple sequence repeats (SSRs), are arrays of short motifs of one to six base pairs in length. These SSR markers are characterized by their hypervariability, abundance, reproducibility, Mendelian inheritance, and codominant nature. The individual identification of botanical samples can help discriminate crime-related samples from other samples and determine their geographic origins. The microsatellites of marijuana (*Cannabis sativa* L., another illicit drug plant) have been extensively studied (5–7), but microsatellites of opium poppies have not been studied at all. The exploitation of microsatellites derived from expressed sequence tag (EST, a unique stretch of DNA within a coding region of a gene from complementary DNA) databases is an enrichment method to reduce costs (8). Therefore, EST-SSR markers of the opium poppy were presently identified and characterized to facilitate the genetic diversity analysis with the aim of application in future forensic cases. cDNA libraries in the National Center for Biotechnology Information (NCBI) were mined for the development of SSR markers of the opium poppy. The candidate markers were used to survey the polymorphisms of 135 individuals. To examine the transferability of the opium poppy EST-SSR markers, the markers were applied to the related species such as five *Papaver* genera species and two *Eschscholzia* genera.

[1]DNA Analysis Sector, Western District Office of National Institute of Scientific Investigation, 111 Daeduk-Ri, Seosam-Myun, Jangsung-Gun, Chonnam 515-822, South Korea.
[2]DNA Analysis Sector, National Institute of Scientific Investigation, 331-1 Shinwol-7 Dong, Yangchun-Ku, Seoul 158-707, South Korea.
[3]Department of Biological Sciences, Chonbuk National University and Jeonju Center of Korea Basic Sciences Institute, 664-14 Dukjin-Dong, Dukjin-Gu, Jeonju City, Jeonbuk 561-756, South Korea.

Received 7 May 2010; and in revised form 11 Aug. 2010; accepted 26 Aug. 2010.

Materials and Methods

Plant Material and Isolation of Total DNA

Samples of *P. somniferum* L. ($n = 135$) were obtained through narcotic seizures from the southwestern province of Korea as part

of a 2-month national government initiative that began in April, 2009, which targeted the illicit cultivation of narcotic plants of the opium poppy. The crack down coincided with the blooming period. One hundred and four samples were seized from Jeonnam, the southern part of Jeolla province, and 31 samples were from Jeonbuk, the northern part of Jeolla. Jeonnam was subdivided into "se," "sw," and "sm" regions and Jeonbuk into "ne" and "nw" regions. Samples were seized from pots and vegetable/flower gardens of violators. Of the 135 samples, The National Police Agency, National Maritime Police Agency, and The Public Prosecutor's Office seized 75, 30, and 30 samples, respectively, from offenders. These agencies submitted requests to the Drug and Toxicology Section of our office for the analysis of drug substances such as morphine and codeine. Each seedling of five *Papaver* genera species (*Papaver bracteatum* [PB], *P. nudicaule* [PN], *P. orientale* [PO], *P. radicatum* [PRA], and *P. rhoeas* [PRH]) and two *Eschscholzia* genera (*Eschscholzia californica* [EC], *E. mexicana* [EM]) were analyzed for the presence of SSR amplicons. The seeds of these seven species were purchased from the Thompson and Morgan seed company (Ipswich, U.K.) and were germinated to obtain seedlings. Genomic DNA was extracted from seedlings, leaf, pod, and stem samples using the Plant DNeasy Mini kit (Qiagen, Valencia, CA) according to the manufacturer's protocol.

EST Mining and EST Primer Design

Expressed sequence tag sequences (n = 20,340) were acquired from NCBI, and potential microsatellite sites were surveyed using SSRIT (9), a microsatellite search tool available from the United States Department of Agricultural–Agriculture Research Station (USDA-ARS) Center for Bioinformatics. Mononucleotides of 10 or more repeats, dinucleotides of six or more repeats, and tri-nucleotides of five or more repeats were extracted (Table 1). Twenty-two primer pairs were designed in the flanking regions of the repeat motifs of the EST sequences using Primer 3 (10)/GeneFisher (11) and were tested as SSR markers. The priming sequences were randomly chosen and focused on di- or tri-nucleotide repeats than other repeats. The locations of microsatellites in coding or noncoding (5′ and 3′ untranslated regions [UTRs]) were assessed using ESTScan (12). Microsatellite markers were designated as "psom X" (Table 2).

PCR Amplification and DNA Genotyping

Polymerase chain reaction (PCR) amplification was carried out in a final volume of 10 μL containing about 25 ng of template DNA, 1 μL of Gold ST*R 10X buffer (Promega, Madison, WI), 0.5 U of AmpliTaq Gold® DNA polymerase (Applied Biosystems, Foster City, CA), and 0.13 μM of each primer pair. Thermal cycling was conducted on GeneAmp® PCR System 9700 (Applied Biosystems) under the following conditions: 95°C for 11 min, 34 cycles of 94°C for 1 min, 56°C for 1 min, 72°C for 1 min, and a final extension at 60°C for 30 min. The primer pairs producing multiple band patterns or null amplicons were not tested in the following steps. When null alleles were observed, two or more amplifications were conducted with higher concentrations of primer pairs and/or DNA polymerase. The 17 SSR amplicons (psom 2, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 16, 17, 18, 19, 20, and 22) were ligated with a pGEM-T vector (Promega) and transformed into JM109 cells by a heat-shock method. Two to four colonies of each SSR amplicon were selected and sequenced with M13 primers for size confirmation using an ABI PRISM 310 Genetic Analyzer (Applied Biosystems). The acquired sequences using 17 primer pairs were registered to NCBI (Table 2).

TABLE 1—*Occurrence of nonredundant SSRs in a set of 20,340 opium poppy ESTs.*

	Number of Repeats												
SSR motif	5	6	7	8	9	10	11	12	13	14	15	>15	Total
A/T						724	324	179	97	84	59	288	1755
C/G						52	23	17	15	6	3	9	125
AC/GT		11	1	1	1				1		1	2	18
AG/CT		38	12	4	2	1	2	3	2	2	1	28	95
AT/TA		43	20	18	14	16	12	6	4	2	3	9	147
GA/TC		38	20	9	8	4	1	2	2		1	10	95
TG/CA		25	5	4	1							1	36
Other NN				2									2
AGA/TCT	24	14	2	1	4		1						46
GAA/TTC	42	12	8	1				1					64
AAC/GTT	6		1			1						2	10
AAG/CTT	26	11	7						1				45
AAT/ATT	15	13	3		2	27		1					61
ACC/GGT	14	2			1								17
ACG/CTG	12	5											17
ACT/ATG	4	2	1		1								8
AGC/CGT	6		1										7
AGG/CCT	5												5
AGT/ATC	15	7											22
CCG/CGG	3	1											4
GAT/GTA	13		2										15
CAG/CAT	19	5	2										26
TAA/TTA	9	5	1	1			1						17
Other NNN	91	19	11	1			1						123
AAAT/ATTT	2												2
AATT/TTAA	1												1
CTTT/TTTC	1												1
TCTT/TTCT	2												2
Other NNNN	1											1	2
NNNNN	3	1										1	5
NNNNNN	3											2	5
NNNNNNNNN								1				1	2
Total													2780

"NN" or "NNN" means all other di- or tri-nucleotides.
EST, expressed sequence tag; SSR, simple sequence repeat.

PCR products with 5′-fluorescein-labeled forward primers and nonlabeled reverse primers were mixed with GeneScan-500 ROX size standards (Applied Biosystems) and electrophoresed in a capillary electrophoresis instrument, the ABI PRISM 310 Genetic Analyzer. The allelic size of each sample was determined by GeneScan software 3.1 and Genotyper 3.7 (Applied Biosystems) comparing with its sequence data. The primer pairs displaying uniform SSR data were not tested in the next steps. Finally, six primer pairs (psom 2, 4, 12, 13, 17, and 22) were used for the acquisition of some statistical data and for the test of cross-species transferability.

All statistical analyses were performed using the population genetic analysis software, GENALEX6, version 6.2 (13), and a dendrogram was graphically displayed by MEGA version 4.0 (14).

Results and Discussion

The SSRIT script identified 2780 microsatellites in 20,340 unigenes (Table 1). The most frequent dimicrosatellite had an AT/TA motif (37%), and the most frequent tri-microsatellite had a GAA/TTC motif (13%). The maximum number of repeats was 60 for the GA motif (NCBI accession number FE966103). The longest motif of repeats was AATAACCCT, and its repeat number was 21 (accession number FG605708). There were more tri-SSRs (487) than di-SSRs (393). Relatively, there were far fewer tetra- and penta-SSRs than tri-SSRs.

TABLE 2—*Characteristics of the* Papaver somniferum *EST-SSR markers.*

Locus	Motif	Location	Primer Sequence (5′-3′)	Size of Cloned Allele (bp)	Allele Size Range (bp)	Na	Ho	He	Referred AC Number	Registered AC Number
psom 2*	(CTT)5C(CTT)6	5′ UTR	ATCCCCTAAACCGTA GGATTACAACCTAATTGAAGC	139	136–139	2	0.00	0.38	FG613716	GU903157
psom 4*	(CA)9	Coding	GCGAAAACCCGCCCTCTAC CCAGAGCCATACTGGGGACA	236/243/255	243–255	4	1.00	0.61	FG613827	GU903158 GU903159 GU903160
psom 5	(TG)8	3′ UTR	GTTAGCAATGGCGGAAGAAC CGCTACCACCTCCGTGTATT	170	–	–	–	–	FG613567	GU903161
psom 6	(TA)15	5′ UTR	TGGGCCCACCTGTATTTGC ACCTCCATGACCACGTGCA	156	–	–	–	–	FG612794	GU903162 GU903163 GU903164
psom 7	(AT)12	Unknown	TGAATCGTCGAGATCAAACG ACATAGGCAGCGAAGTTGGT	209	–	–	–	–	FG610974	GU903165
psom 8	(GA)8	5′ UTR	CGAACCACGGATTACCCTAA TCCCTCCGTTTCACAAATTC	221	–	–	–	–	FG612134	GU903166
psom 9	(AGA)6	Coding	CAGCAGAAGTCGTGCAGAAG TTCTTTGCGACAGTCAGTGG	249	–	–	–	–	FG607359	GU903167 GU903168
psom 11	(TCT)9	Unknown	GCAGTGGAATGGGTTGAGTT TGGACCAAGTAGGGATCCAG	180/198	–	–	–	–	FG610874	GU903169 GU903170 GU903171
psom 12*	(TTC)8	Coding	CCCTGCTGCTTTCAAATCTC ATGGAAGAAAAGGTGCCTGA	234	234–237	2	0.00	0.42	FG610824	GU903172 GU903173
psom 13*	(CAG)7	Unknown	ACCCCCACAACTTTGTCGTA CGAGTGTTTGGGAACCTGAT	151	145–151	2	0.00	0.49	FG606729	GU903174 GU903175
psom 14	(TC)8	3′ UTR	CTACATCCACGGGGTTTGTT CCAATCTCCACCAAGAGGAA	174/180	–	–	–	–	FG606181	GU903176 GU903177 GU903178 GU903179
psom 16	(GAA)8	Coding	AAAATGGGAACAGCCATCAG TCTGGGCCTTCACCACTTAC	223	–	–	–	–	FE967887	GU903180
psom 17*	(ATT)10	5′ UTR	CCCAACAATTTGGTGCAGTA CCCGAGGTGAAACTTCTCTG	150/165	159–177	5	0.00	0.47	FE967526	GU903181 GU903182 GU903183
psom 18	(CAA)5	Coding	GCGTTTACTCAGGTGGAAGC AGCTTGGGTCTTTTCCCATT	236	–	–	–	–	FE967151	GU903184
psom 19	(CCT)5(CAC)6	Coding	ACCACCACCTACTGCTTCTC GCTCACACATGGGCTGCT	154	–	–	–	–	FE967394	GU903185
psom 20	(CCT)5(CAC)7	Coding	TACCCTCCACCACCTACTGC CACATGGGCTGCTTTGATA	207/213	–	–	–	–	FE966396	GU903186 GU903187
psom 22*	(TAA)6	Coding	GTAGGGAGGTGTGCTCCAGT TCTTTGGCTCCGCCTTGGA	312/315	315–318	2	0.00	0.49	CO267815	GU903188 GU903189 GU903190

*Loci that were used in this statistical study. bp, base pair; Na, allele number; Ho, represents observed heterozygosity; He, represents expected heterozygosity; AC number, means accession number in NCBI; EST, expressed sequence tag; SSR, simple sequence repeat; UTR, untranslated regions.

Twenty-two primer pairs were designed for di- and tri-SSRs. Of these 22 primer pairs, 11 were from coding regions, four were from 5′ UTR regions, three were from 3′ UTR regions, and four were from unknown regions. Seventeen of the primer pairs successfully amplified opium poppy genomic DNA, and their amplicons were cloned for sequence acquisition (Table 2). The remaining five primer pairs showed nonamplicon (psom 3 and 5) or odd binding patterns (psom 1, 10, and 21). Although four primer pairs (psom 5, 7, 11, and 14) of 17 produced successful amplicons, their genotyping results could not be included in further analyses because of their multiple (≥3) allele peaks. The amplified product with seven primer pairs (psom 6, 8, 9, 16, 18, 19, and 20) of 13 did not exhibit any variability against the first 32 individuals of *P. somniferum* L. The remaining six (psom 2, 4, 12, 13, 17, and 22) microsatellite loci of 13 were selected, and their SSR polymorphisms were surveyed against 135 individuals from a population of *P. somniferum* L.

Using six SSR markers (psom 2, 4, 12, 13, 17, and 22), an average of 2.83 ± 0.54 (mean ± SE, range 2–5) alleles per locus with a mean expected heterozygosity of 0.48 ± 0.03 (mean ± SE, range 0.38–0.61) was detected (Table 2). Four SSR markers (psom 2, 12, 13, and 22) exhibited only two alleles, psom 4 marker showed four alleles, and psom 17 marker showed five alleles (Fig. 1). The low genetic differentiation power of the EST-SSR markers was because of the fact that the EST was a short sub-sequence of a transcribed cDNA sequence essential for a life-sustaining mechanism (some

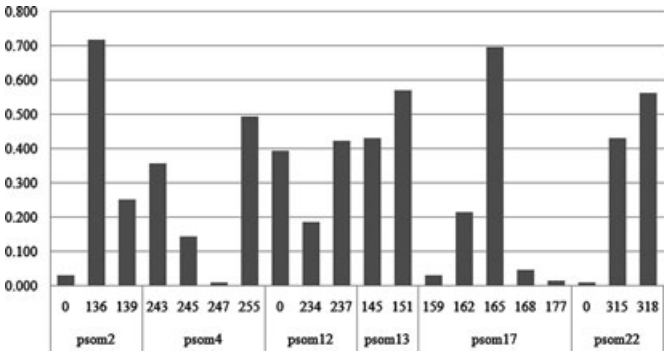

FIG. 1—*Allele frequencies of the six loci in population from 135 opium poppies. "0" represents missing data at the loci psom 2, 12, and 22.* x*-axis shows six loci and* y*-axis shows allele frequencies.

variations in cDNA sequences may be related to natural selection [15]). EST-derived SSRs are significantly less polymorphic than those derived from genomic regions (8). SSR and short tandem repeat were typically observed in the noncoding intron regions.

Null alleles were detected in 53 individuals for the psom 12 primer, four individuals for psom 2, and one individual for psom 22. Null allele was included in the population statistic analysis as zero (Fig. 1). Point mutations in the primer annealing sites may have led to the occurrence of null alleles, where the microsatellites fail to amplify in PCR assays (16,17).

Cross-species transferability was detected in two markers, psom 4 and psom 17. Five *Papaver* genera species (PB, PN, PO, PRA, and PRH) and two *Eschscholzia* genera (EC and EM) exhibited amplicons with psom 4 primer pairs. Three species of *Papaver* genera species (PB, PO, and PRH) exhibited amplicons with psom 17 primer pairs. The primer pairs of psom 4 and psom 17 might be used as *Papaver* genus-specific marker for the identification of opium, oriental, and corn poppies from other poppy species that share similar external characteristics. It might be significant that the population genetic studies of the poppy species using EST-SSRs are being conducted throughout the world.

Six loci identified 29 unique profiles for the 135 individuals. There were 32 individuals with the most frequent types and 23, 15, and 10 individuals with the next three most frequent types, respectively. The EST-SSRs exhibited small numbers of alleles per locus (2–5) and low heterozygosity (0.38–0.61) (Table 2). This phenomenon was observed in another study using random-amplified polymorphic DNA and inter-SSR amplification of 24 opium poppy germplasm (18). A neighbor-joining dendrogram of circular form was acquired from the 135 seized opium poppies based on the genetic distances of alleles for the six EST-SSR markers (Fig. 2). The dendrogram was comprised of two major groups. The left and right major groups could be divided into two and three subgroups, respectively. But the specimens did not cluster together into a single unit on the basis of geography, which was observed previously in an amplified fragment length polymorphism study of opium poppy (19).

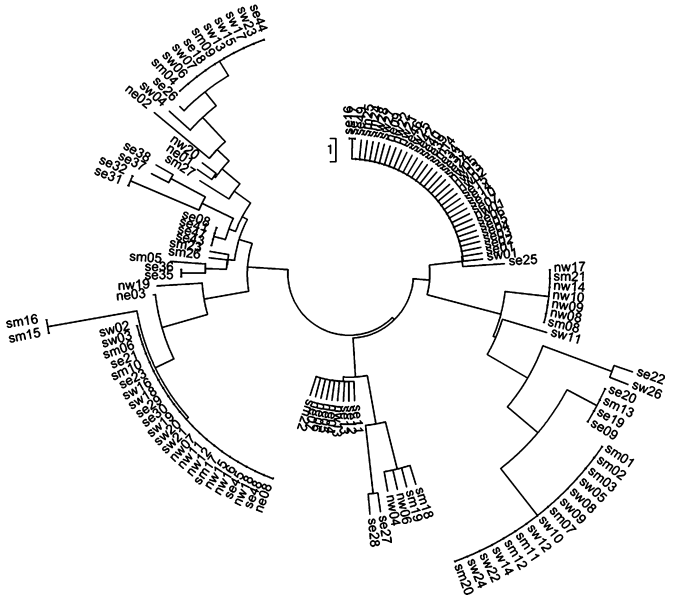

FIG. 2—*Neighbor-joining dendrogram of circular form for 135 individual opium poppies based on the allelic distances for six expressed sequence tag–simple sequence repeat markers.*

In Korea, opium has been commonly used for household pain and paregoric medicine for many generations. Many of the seized samples of opium poppy were from remote mountainous or coastal regions in outlying rural districts. The restricted cultivation area and the limited gene pool of the opium poppy might be the cause of its low genetic diversity. Surveys of SSR frequencies of the representative samples from other provinces in South Korea should be undertaken. Furthermore, studies of other variable markers based on genomic DNA could enhance the discrimination power of individual identification of opium poppy.

Acknowledgments

This work using opium poppy was authorized by the Korea Food and Drug Administration. The authors kindly thank Hee-Jin Yang, Il-Woong Seol, and Ju-Seon Lee of the Drug & Toxicology Section in the Western District Office of the National Institute of Scientific Investigation for the supply of opium poppy specimens.

References

1. Gjerde H, Fongen U, Gundersen H, Christophersen AS. Evaluation of a method for simultaneous quantification of codeine, ethylmorphine and morphine in blood. Forensic Sci Int 1991;51(1):105–10.
2. Geier A, Bergemann D, von Meyer L. Evaluation of a solid-phase extraction procedure for the simultaneous determination of morphine, 6-monoacetylmorphine, codeine and dihydrocodeine in plasma and whole blood by GC/MS. Int J Legal Med 1996;109(2):80–3.
3. Vincent PG, Engelke BF. High pressure liquid chromatographic determination of the five major alkaloids in *Papaver somniferum* L. and thebaine in *Papaver bracteatum* Lindl. capsular tissue. J Assoc Off Anal Chem 1979;62(2):310–4.
4. Kapoor LD. Opium poppy: botany, chemistry, and pharmacology. New York, NY: Haworth Press, 1995.
5. Alghanim HJ, Almirall JR. Development of microsatellite markers in *Cannabis sativa* for DNA typing and genetic relatedness analyses. Anal Bioanal Chem 2003;376(8):1225–33.
6. Gilmore S, Peakall R, Robertson J. Short tandem repeat (STR) DNA markers are hypervariable and informative in *Cannabis sativa*: implications for forensic investigations. Forensic Sci Int 2003;131(1): 65–74.
7. Mendoza MA, Mills DK, Lata H, Chandra S, ElSohly MA, Almirall JR. Genetic individualization of *Cannabis sativa* by a short tandem repeat multiplex system. Anal Bioanal Chem 2009;393(2):719–26.
8. Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 2003;106(3):411–22.
9. Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 2001;11:1441–52.
10. Rozen S, Skaletsky HJ. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, editors. Bioinformatics methods and protocols: methods in molecular biology. Totowa, Canada: Humana Press, 2000;365–86.
11. Giegerich R, Meyer F, Schleiermacher C. GeneFisher-software support for the detection of postulated genes. Proc Int Conf Intell Syst Mol Biol 1996;4:68–77.
12. Iseli C, Jongeneel CV, Bucher P. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proc Int Conf Intell Syst Mol Biol 1999;138–48.
13. Peakall R, Smouse PE. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 2007;6(1):288–95.
14. Kumar S, Dudley J, Nei M, Tamura K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform 2008;9:299–306.
15. Kado T, Yoshimaru H, Tsumura Y, Tachida H. DNA variation in a conifer, *Cryptomeria japonica* (Cupressaceae sensu lato). Genetics 2003; 164(4):1547–59.

16. Jarne P, Lagoda PJL. Microsatellites, from molecules to populations and back. Trends Ecol Evol 1996;11:424–9.
17. Dakin EE, Avise JC. Microsatellite null alleles in parentage analysis. Heredity 2004;93(5):504–9.
18. Acharya HS, Sharma V. Molecular characterization of opium poppy germplasm. Am J Infect Dis 2009;5(2):155–60.
19. Saunders JA, Pedroni MJ, Penrose LDJ, Fist AJ. AFLP analysis of opium poppy. Crop Sci 2001;41:1596–601.

Additional information and reprint requests:
Moon Sik Yang, Ph.D.
Department of Biological Sciences
Chonbuk National University and Jeonju Center of Korea Basic Sciences Institute
664-14 Dukjin-Dong, Dukjin-Gu, Jeonju City, Jeonbuk 561-756
South Korea
E-mail: mskyang@chonbuk.ac.kr