Genetics and Molecular Research

Molecular identification and phylogenetic analysis of important medicinal plant species in genus *Paeonia* based on rDNA-ITS, *mat*K, and *rbc*L DNA barcode sequences

W.J. Kim, Y. Ji, G. Choi, Y.M. Kang, S. Yang and B.C. Moon

K-herb Research Center, Korea Institute of Oriental Medicine, Yuseong-daero, Yuseong-gu, Daejeon, Republic of Korea

Corresponding author: BC Moon
E-mail: bcmoon@kiom.re.kr

Genet. Mol. Res. 15 (3): gmr.15038472
Received January 20, 2016
Accepted March 11, 2016
Published August 5, 2016
DOI http://dx.doi.org/10.4238/gmr.15038472

ABSTRACT. This study was performed to identify and analyze the phylogenetic relationship among four herbaceous species of the genus *Paeonia*, *P. lactiflora*, *P. japonica*, *P. veitchii*, and *P. suffruticosa*, using DNA barcodes. These four species, which are commonly used in traditional medicine as Paeoniae Radix and Moutan Radicis Cortex, are pharmaceutically defined in different ways in the national pharmacopoeias in Korea, Japan, and China. To authenticate the different species used in these medicines, we evaluated rDNA-internal transcribed spacers (ITS), *mat*K and *rbc*L regions, which provide information capable of effectively distinguishing each species from one another. Seventeen samples were collected from different geographic regions in Korea and China, and DNA barcode regions were amplified using universal primers. Comparative analyses of these DNA barcode

sequences revealed species-specific nucleotide sequences capable of discriminating the four *Paeonia* species. Among the entire sequences of three barcodes, marker nucleotides were identified at three positions in *P. lactiflora*, eleven in *P. japonica*, five in *P. veitchii*, and 25 in *P. suffruticosa*. Phylogenetic analyses also revealed four distinct clusters showing homogeneous clades with high resolution at the species level. The results demonstrate that the analysis of these three DNA barcode sequences is a reliable method for identifying the four *Paeonia* species and can be used to authenticate Paeoniae Radix and Moutan Radicis Cortex at the species level. Furthermore, based on the assessment of amplicon sizes, inter/intra-specific distances, marker nucleotides, and phylogenetic analysis, rDNA-ITS was the most suitable DNA barcode for identification of these species.

Key words: DNA barcode; *Paeonia* species; rDNA-ITS; *mat*K; *rbc*L; Species discrimination

INTRODUCTION

In Korean herbal medicine, approximately 540 species of medicinal plants have been traditionally applied to the treatment of many diseases and symptoms. The *Paeonia* genus includes two important traditional herbal medicines, Paeoniae Radix and Moutan Radicis Cortex. Moutan Radicis Cortex is exclusively made from *P. suffruticosa*; however, Paeoniae Radix is described differently in the Korean, Chinese, and Japanese pharmacopoeia (Korea Institute of Oriental Medicine, 2015). In addition, in the Pharmacopoeia of the People's Republic of China, *P. lactiflora* is used in two distinct herbal medicines, Paeniae Radix Rubra and Paeoniae Radix Alba (Korea Institute of Oriental Medicine, 2015). Paeoniae Radix Alba and *P. japonica* are both referred to as "Baekjak" and consequently, *P. japonica* and *P. lactiflora* may be confused due to the similar pronunciations of these homonyms.

In prescriptions for traditional medicine, Paeoniae Radix and Moutan Radicis Cortex are used to treat different symptoms. Paeoniae Radix is commonly used to treat blood extravasation, relieve pain, clear fever, and protect the liver (Li et al., 2016). In contrast, Moutan Radicis Cortex is used to activate the blood and treat blood extravasation in the heart, liver, and kidney (Kim et al., 2007; Korea Institute of Oriental Medicine, 2015). Therefore, in order to prevent inaccurate or inauthentic prescriptions that include these herbal medicines, it is very important to distinguish among the four *Paeonia* species.

Plant species and herbal medicines are identified using a variety of methods, including analyses of morphologic characteristics, chemical compositions, and molecular genetic variations (Jiang et al., 2014; Moon et al., 2015; Zhao et al., 2015). For successful morphological identification, a pharmacognostic expert must have extensive professional experience that is acquired from evaluating large numbers of medicinal herbs. In addition, an objectively assessable classification key must be constructed based on internal or external morphological traits, such as the starch grain of parenchyma cells, shape of phloem or xylem, crystal of the cortex, and epidermis or cross-section of the dried medicinal herb (Mahadani and Ghosh, 2013; Chen et al., 2014). Chemical analysis requires costly techniques, including high-performance liquid chromatography, UV/Vis spectrophotometry, and near-infrared

spectroscopy (Jiang et al., 2014; Wang et al., 2015). Moreover, the results of chemical analysis can vary depending on environmental factors and physiological conditions, and the necessary supplies can be extremely expensive (Joshi et al., 2004; Jiang et al., 2014).

Molecular markers represent a potentially cheaper and more accurate alternative to chemical analyses. Because DNA markers are based on unique nucleotide sequences and are not affected by environmental factors or physiological conditions, their analysis can provide rapid and accurate monitoring of multiple types of samples including fresh plants, dried medicinal herbs, and processed foods (Joshi et al., 2004; Ma et al., 2014; Vassou et al., 2015). Among the diverse molecular genetic tools available for identifying medicinal plant species, analysis of DNA barcodes based on short DNA sequences in the nuclear or organellar genomes is commonly used for efficient and accurate species identification (Jiang et al., 2014; Zhao et al., 2015). The chloroplast and nuclear genomes contain multiple candidate DNA barcode regions with high substitution rates. Among these regions, rDNA-internal transcribed spacers (ITS), *rbc*L, *mat*K, *psb*A-*trn*H, *trn*L-F, and *trn*H have been used to identify plant species and analyze phylogenetic relationships (Tripathi et al., 2013; Chen et al., 2014; Aziz et al., 2015; Vassou et al., 2015). The *rbc*L region is considered to be a universal DNA barcode region, but it is less useful for species identification due to its slow rate of molecular evolution (Tripathi et al., 2013; Ma et al., 2014). Another universal DNA barcode region, *mat*K, exhibits higher rates of sequence variation and is therefore, superior for purposes of species discrimination. The Plant Working Group of the Consortium for the Barcode of Life (CBOL) recommends using a combination of *rbc*L and *mat*K regions as a universal DNA barcode for Plantae (CBOL Plant Working Group, 2009). Alternatively, the rDNA-ITS is suitable as a barcode region because it can very reliably identify species in diverse plant genera (Tripathi et al., 2013; Vassou et al., 2015; Zhao et al., 2015). Consequently, rDNA-ITS sequences have the potential to distinguish between medicinal plants and their close relatives. Recently, these DNA barcode regions have frequently been applied to diverse herbal medicines to authenticate medicinal plant species and analyze the molecular phylogeny of medicinal herbs (Kim et al., 2013; Ma et al., 2014; Aziz et al., 2015; Zheng et al., 2015). In this study, we comparatively analyzed the three DNA barcode regions, rDNA-ITS, *mat*K, and *rbc*L, to evaluate their ability to identify the four, potentially confused medicinal *Paeonia* species. We identified species-specific single nucleotide polymorphisms (SNPs, i.e., marker nucleotides) that are sufficient to discriminate among these species, and confirmed that these DNA barcodes are suitable for the taxonomic identification and classification of *Paeonia* samples at the species level.

MATERIAL AND METHODS

Plant material

Seventeen samples of *P. lactiflora*, *P. japonica*, *P. veitchii*, or *P. suffruticosa* were used as plant material in this study (Table 1). Samples collected from different geographical regions of the Republic of Korea and China were quick-frozen in liquid nitrogen and then stored at -80°C in a deep freezer. Independent individuals of each species were identified by the Classification and Identification Committee of the Korea Institute of Oriental Medicine (KIOM), which is comprised of experts in the fields of plant taxonomy, botany, pharmacognosy, and herbology. These plant materials were provided accession numbers and specimens were preserved in the Herbarium of the KIOM.

Table 1. Properties of taxa used in this study.

Abbreviation	Scientific name	Herbal name and origin			Location	Voucher	GenBank accession No.		
		KP 11†	ChP 2010‡	JP XVI⁋			rDNA-ITS	*mat*K	*rbc*L
PL_JS	*Paeonia lactiflora* Pall.	Paeoniae Radix	Paeoniae Radix Alba/ Paeniae Radix Rubra	Paeoniae Radix	Jeongseon-eup, Jeongseon-gun, Gangwon-do, Korea	KIOM2013KR02-50	KT944679	KT944696	KT944713
PL_GC					Gabuk-myeon, Geochang-gun, Gyeongsangnam-do, Korea	KIOM2013KR04-45	KT944680	KT944697	KT944714
PL_HC					Gaya-myeon, Hapcheon-gun, Gyeongsangnam-do, Korea	KIOM2013KR05-07	KT944681	KT944698	KT944715
PL_HL					Helong, Yanbian Korean Autonomous Prefecture, Jilin, China	KIOM2013CN01-21	KT944682	KT944699	KT944716
PL_YJ					Yanji, Yanbian Korean Autonomous Prefecture, Jilin, China	KIOM2013CN01-46	KT944683	KT944700	KT944717
PJ_GR	*Paeonia japonica* (Makino) Miyabe & Takeda	Paeoniae Radix	-¶	-¶	Sandong-myeon, Gurye-gun, Jeollanam-do, Korea	KIOM2007KR01-02	KT944684	KT944701	KT944718
PJ_MJ					Anseong-myeon, Muju-gun, Jeollabuk-do, Korea	KIOM2007KR01-03	KT944685	KT944702	KT944719
PJ_IJ					Girin-myeon, Inje-gun, Gangwon-do, Korea	KIOM2007KR02-13	KT944686	KT944703	KT944720
PJ_HY					Anui-myeon, Hamyang-gun, Gyeongsangnam-do, Korea	KIOM2012KR07-04	KT944687	KT944704	KT944721
PJ_SGP					Saekdal-dong, Seogwipo-si, Jeju-do, Korea	KIOM2013KR03-47	KT944688	KT944705	KT944722
PV_SG1	*Paeonia veitchii* Lynch	Paeoniae Radix	Paeniae Radix Rubra	-¶	Siguniang Mt., Xiaojin, Ngawa, Sichuan, China	KIOM2012CN02-10	KT944689	KT944706	KT944723
PV_SG2					Siguniang Mt., Xiaojin, Ngawa, Sichuan, China	KIOM2012CN02-11	KT944690	KT944707	KT944724
PV_SG3					Siguniang Mt., Xiaojin, Ngawa, Sichuan, China	KIOM2012CN02-12	KT944691	KT944708	KT944725
PV_SG4					Siguniang Mt., Xiaojin, Ngawa, Sichuan, China	KIOM2012CN02-13	KT944692	KT944709	KT944726
PS_AD1	*Paeonia suffruticosa* Andrews	Moutan Radicis Cortex	Moutan Cortex	Moutan Cortex	Pungsan-eup, Andong-si, Gyeongsangbuk-do, Korea	KIOM2013KR13-49	KT944693	KT944710	KT944727
PS_AD2					Pungsan-eup, Andong-si, Gyeongsangbuk-do, Korea	KIOM2013KR13-50	KT944694	KT944711	KT944728
PS_AD3					Pungsan-eup, Andong-si, Gyeongsangbuk-do, Korea	KIOM2014KR01-13	KT944695	KT944712	KT944729

†: Korean Pharmacopoeia, 11th edition. ‡: Pharmacopoeia of the People's Republic of China, 2010 edition. ⁋: Japanese Pharmacopoeia, 16th edition. ¶: There is no appropriate official name in the corresponding national pharmacopoeia.

Genomic DNA extraction

Genomic DNA was extracted from approximately 100 mg fresh leaves using a Precellys™ Grinder (Bertin Technologies, Montigny-le-Bretonneux, France) and the DNeasy® Plant Mini Kit (Qiagen, Valencia, CA, USA). Quality and quantity of genomic DNA were monitored using UV/Vis spectrophotometry (ND-1000, NanoDrop, Wilmington, DE, USA) and resolution was determined using 1.5% agarose gel electrophoresis with a 100-bp DNA ladder (Solgent, Daejeon, Korea). Each genomic DNA sample was diluted to approximately 15 ng/μL in TE (pH 8.0) for use as a PCR template.

PCR amplification of DNA barcoding regions and sequencing

The DNA barcoding regions, rDNA-ITS, *mat*K, and *rbc*L were amplified using universal primers for vascular plants: ITS1 (5'- TCC GTA GGT GAA CCT GCG G-3') and ITS4 (5'- TCC TCC GCT TAT TGA TAT GC-3') for the entire ITS region; *mat*K AF (5'- CTA TAT CCA CTT ATC TTT CAG GAG T-3') and *mat*K 8R (5'- AAA GTT CTA GCA CAA GAA AGT CGA-3') for a ~1.3-kb fragment of the *mat*K gene; and *rbc*L F (5'-ATG TCA CCA CAA ACA GAA ACT AAA GC-3') and *rbc*L R (5'-TCC TTT TAG TAA AAG ATT GGG CGG AG-3') for a ~1.6-kb fragment of the *rbc*L gene (White et al., 1990; Olmstead and Reeves, 1995; Kato et al., 1998). PCR mixtures (40 μL total volume) consisted of approximately 15 ng genomic DNA, 20 pM each forward and reverse primers, Solg™ 2X Taq PCR Smart-Mix I (Solgent), and distilled, deionized water. The PCR conditions were as follows: initial denaturation for 2 min at 95°C; 35 cycles of 1 min denaturation at 95°C, 40 s annealing at 53°C, and 1 min extension at 72°C; and final extension for 5 min at 72°C. The amplified PCR products were resolved by electrophoresis on 1.5% agarose gels with a 100-bp DNA ladder (Solgent). Specific amplified bands, approximately 750 bp for rDNA-ITS, 1.3 kb for *mat*K, and 1.5 kb for *rbc*L were purified from the gels using the Gel Extraction Kit (Qiagen) and sub-cloned into pGEM®-T Easy Vector (Promega, Madison, WI, USA). The ligation reaction was prepared as follows: 4.5 μL DNA; 0.5 μL pGEM®-T Easy Vector; and 5 μL Ligation Mix (DNA Ligation Kit <Mighty Mix>, TaKaRa Bio Inc., Otsu, Shiga, Japan). The ligation mixture was incubated at room temperature for 30 min, and then 5 μL ligation mixture was used to transform competent cells (*Escherichia coli* strain JM109). Recombinant white colonies were picked and sequenced. Both strands of the insert DNA were sequenced by dideoxynucleotide chain termination on an automatic DNA sequence analyzer (ABI3730, Applied Biosystems Inc., Foster City, CA, USA). Each DNA barcode region was sequenced a minimum of three times per sample. The resultant nucleotide sequences of the rDNA-ITS, *mat*K, and *rbc*L genes for all 17 samples were registered in NCBI GenBank (Table 1).

Sequence analysis and genetic relationship analysis

Approximately 800-bp sequences obtained from bi-directional sequencing with T7 and SP6 promotor were assembled and edited using the BioEdit version 7.2.5 (Hall, 1999) software. The contigs were aligned to analyze the intra- and inter-species variations in the sequence. For the analysis of sequence identity and evolution, inter- or intra-species genetic distances were calculated using the Kimura-2-parameter (K2P) model in the MEGA6 (Molecular Evolutionary Genetics Analysis Version 6.0) program (Tamura et al., 2013, http://

www.megasoftware.net). The phylogenetic analysis based on the rDNA-ITS, *mat*K, and *rbc*L sequences was performed using MEGA version 6.06 (Kimura, 1980; Tamura et al., 2013) software. The phylogenetic tree was constructed using the neighbor-joining (NJ) method with the K2P model, pairwise deletion for gaps/missing data treatment, and 1000 replications for bootstrap. *Hamamelis japonica* was chosen as an out-group control. The resulting NCBI GenBank accession numbers are GU576652 for rDNA-ITS, AF198091 for *mat*K, and AY263940 for *rbc*L.

RESULTS

PCR amplification and genetic diversity of DNA barcoding region

PCR was performed on all 17 *Paeonia* samples. Amplification of the three DNA barcode regions, rDNA-ITS, *mat*K, and *rbc*L, with universal primers yielded 100% PCR efficiency. Thus, the universal primers, ITS1/ITS4, *mat*K AF/*mat*K 8R, and *rbc*L F/*rbc*L R, are suitable for use in the genus *Paeonia*. Amplicons were sequenced after sub-cloning and the final lengths of the sequences were 724 bp for rDNA-ITS; 1511-1530 bp for *rbc*L; and 1238 bp for *mat*K (Table 2). Genetic divergence within or between the species was calculated using the DNA barcode sequences (**Tables S1, S2** and **S3**). The intra-specific distances ranged from 0.18% to 0.95% in rDNA-ITS, 0.00 to 0.05% in *mat*K, and 0.00 to 0.07% in *rbc*L. In rDNA-ITS, *P. veitchii* exhibited the highest variation rate among the four species (0.95%), whereas *mat*K and *rbc*L differed very little relative to the other three species (0.00 and 0.03%, respectively). Similar to the intra-specific distance for the three DNA barcodes in *P. veitchii*, the variability of the rDNA-ITS sequence was also greater than those for *mat*K and *rbc*L in the other three *Paeonia* species. The average inter-specific distance of the rDNA-ITS region was highest in *P. suffruticosa*, followed by *P. veitchii*, *P. japonica*, and *P. lactiflora*; the variability of *mat*K was highest in *P. suffruticosa*, followed by *P. japonica*, *P. veitchii*, and *P. lactiflora*; and the variability of *rbc*L was highest in *P. suffruticosa*, followed by *P. veitchii*, *P. lactiflora*, and *P. japonica*. The inter-species distances ranged from 2.35 to 2.73% in rDNA-ITS, 0.49 to 1.22% in *mat*K, and 0.30 to 0.40% in *rbc*L. *P. suffruticosa* had the highest inter-specific distance in all three DNA barcode regions, and *P. lactiflora* had the lowest distance in both rDNA-ITS and *mat*K. These results suggest that the traditional DNA barcodes, rDNA-ITS, *mat*K, and *rbc*L, could discriminate the original plant species between Paeoniae Radix and Moutan Radicis Cortex and rDNA-ITS sequence is the most suitable DNA marker for distinguishing between four *Paeonia* species.

Sequence analysis and species discrimination

Species-specific substitutions and indels, i.e., marker nucleotides that are sufficient to authenticate the individual species, were identified from a comparative analysis of the entire nucleotide sequences of the four *Paeonia* species (Tables 2-5). From the comparative alignments of rDNA-ITS, we obtained species-specific substitutions at one position for *P. lactiflora*, four for *P. japonica*, two for *P. veitchii*, and nine for *P. suffruticosa* (Table 3). We defined these species-specific substitutions as marker nucleotides because they were crucial for identifying each species. For example, when all of the 17 rDNA-ITS sequences of four *Paeonia* species were aligned against the *P. lactiflora* PL_JS sample reference sequence,

cytosine (C) 83 was substituted for guanine (G) only in *P. suffruticosa* (Table 3). This unique species-specific substitution is enough to discriminate *P. suffruticosa* from the three other *Paeonia* species. In *mat*K, species-species marker nucleotides were identified at five positions in *P. japonica*, one in *P. veitchii*, and 12 in *P. suffruticosa*, but not in *P. lactiflora* (Table 4).

Table 2. Statistics for three DNA barcodes used in this study.

DNA barcode region	Scientific name	Amplicon length (bp)	Intra-specific distance	Inter-specific distance	Marker nucleotide site
rDNA-ITS	*P. lactiflora*	742	0.0057 ± 0.0033	0.0235 ± 0.0055	1
	P. japonica	742	0.0024 ± 0.0017	0.0243 ± 0.0034	4
	P. veitchii	742	0.0095 ± 0.0038	0.0249 ± 0.0069	2
	P. suffruticosa	742	0.0018 ± 0.0008	0.0273 ± 0.0060	9
*mat*K	*P. lactiflora*	1238	0.0005 ± 0.0004	0.0049 ± 0.0037	0
	P. japonica	1238	0.0003 ± 0.0004	0.0072 ± 0.0043	5
	P. veitchii	1238	0.0000 ± 0.0000	0.0049 ± 0.0039	1
	P. suffruticosa	1238	0.0005 ± 0.0005	0.0122 ± 0.0019	12
*rbc*L	*P. lactiflora*	1530	0.0003 ± 0.0003	0.0031 ± 0.0007	2
	P. japonica	1511	0.0007 ± 0.0005	0.0030 ± 0.0063	2
	P. veitchii	1511	0.0003 ± 0.0004	0.0031 ± 0.0007	2
	P. suffruticosa	1511	0.0000 ± 0.0000	0.0040 ± 0.0003	4

Table 3. rDNA-ITS sequence variation among the four *Paeonia* species.

	Aligned position														
	79	83	99	104	113	118	123	161	168	169	544	559	651	658	678
P. lactiflora	A	C	**T**	T	T	G	G	A	A	G	A	C	G	A	A
P. japonica	**G**	·	C	·	·	·	·	**C**	**T**	**C**	·	·	·	·	·
P. veitchii	·	·	C	**C**	·	·	·	·	·	·	·	**T**	·	·	·
P. suffruticosa	**C**	**G**	C	·	**C**	**A**	**A**	·	·	·	**G**	·	**A**	**G**	**T**

Dots (·) indicate nucleotides identical to those of *P. lactiflora*. Underlined bold characters represent species-specific substitutions.

Table 4. *mat*K sequence variation among the four *Paeonia* species.

	Aligned position																		
	48	134	138	139	348	385	533	537	540	588	625	772	823	903	935	983	1027	1040
P. lactiflora	G	T	A	C	T	G	A	A	C	C	C	C	G	C	G	C	T	C
P. japonica	**A**	·	·	**T**	·	·	·	**G**	·	**T**	**T**	·	·	·	·	·	·	·
P. veitchii	·	·	·	·	**G**	·	·	·	·	·	·	·	·	·	·	·	·	·
P. suffruticosa	·	**G**	**G**	·	·	**T**	**C**	·	**A**	·	·	**T**	**A**	**T**	**T**	**T**	**C**	**T**

Dots (·) indicate nucleotides identical to those of *P. lactiflora*. Underlined bold characters represent species-specific substitutions.

Table 5. *rbc*L sequence variation among the four *Paeonia* species.

	Aligned position									
	507	673	723	836	873	1058	1308	1338	1345	1416–1434
P. lactiflora	A	**C**	T	C	T	**A**	T	T	T	----------
P. japonica	·	A	·	·	·	T	·	·	**A**	**GGATACAATTTGAAGCGAT**
P. veitchii	**G**	A	·	·	**C**	T	·	·	·	----------
P. suffruticosa	·	A	**C**	**G**	·	T	**C**	**C**	·	----------

Dots (·) indicate nucleotides identical to those of *P. lactiflora*, and dashes (-) represent nucleotide deletions at the aligned nucleotide positions. Underlined bold characters and dashes (-) represent species-specific substitutions and indels, respectively.

From the analysis of *rbc*L sequences, we also obtained species-specific marker nucleotides at two positions in *P. lactiflora*, one position in *P. japonica*, two positions in *P. veitchii*, and four positions in *P. suffruticosa* (Table 5). In addition to these species-specific substitutions, a 19-bp *P. japonica*-specific indel region that was applicable to the marker nucleotides, was the only such sequence in the three barcode regions, and was observed in *rbc*L (Table 5). Although *P. lactiflora*-specific marker nucleotides were not identified in *mat*K, the combination of *mat*K+*rbc*L, which is a universal plant DNA barcode recommended by CBOL, provided enough information to authenticate each of the four *Paeonia* species. The results obtained using the sequence differences in the three DNA barcodes revealed that *P. suffruticosa*, which had the highest inter-specific distance, i.e., the highest genetic variation relative to the other three *Paeonia* species (Table 2), had the highest number of marker nucleotides. Thus, inter-specific distance positively correlates with the presence of marker nucleotides.

Phylogenetic analyses

To evaluate the phylogenetic relationships among the four herbaceous *Paeonia* species, phylogenetic trees were constructed by applying the NJ method to the entire sequences of rDNA-ITS, *mat*K, and *rbc*L. All 17 individual samples were clustered into four groups constituting homogeneous clades within each species that reflected the intra-specific variation in all DNA barcodes (Figures 1-3). As shown in the results of the phylogenetic tree, when all three DNA barcode sequences were employed, *P. lactiflora* was closer genetically to *P. veitchii* than to the other two *Paeonia* species, while *P. suffruticosa* was more distant to the other three *Paeonia* species. Bootstrap values clustered at the species level ranged from 81 to 99% in rDNA-ITS, 60 to 99% in *mat*K, and 53 to 98% in *rbc*L (Figures 1-3). Three of the *Paeonia* species had bootstrap values greater than 60%; the exception was the *P. suffruticosa rbc*L gene, which clustered at 53%. In particular, the rDNA-ITS allowed a clearer classification of the four species than either *mat*K or *rbc*L. To distinguish species based on the phylogenetic tree, the minimum bootstrap values were 81% in rDNA-ITS, 60% in *mat*K, and 53% in *rbc*L. Therefore, the rDNA-ITS barcode sequence provided higher resolution for the identification of clusters that constitute homogeneous clades within each *Paeonia* species. From these phylogenetic analyses, we confirm that the identification of the four *Paeonia* species can be achieved using sequence variability in their rDNA-ITS, *mat*K, and *rbc*L sequences.

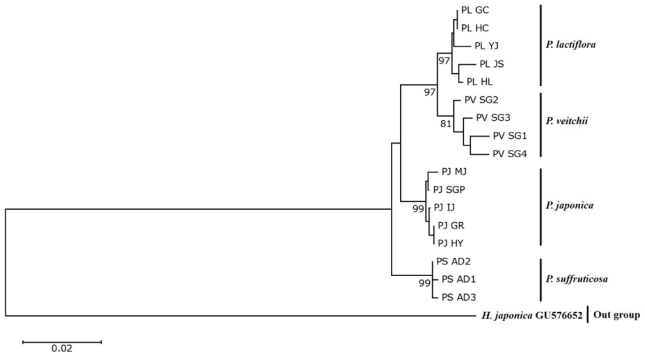

Figure 1. Phylogenetic tree with 1000 bootstrap replicates based on rDNA-ITS sequences from four *Paeonia* species and *Hamamelis japonica*. NJ bootstrap values greater than 60% are shown above the branches.

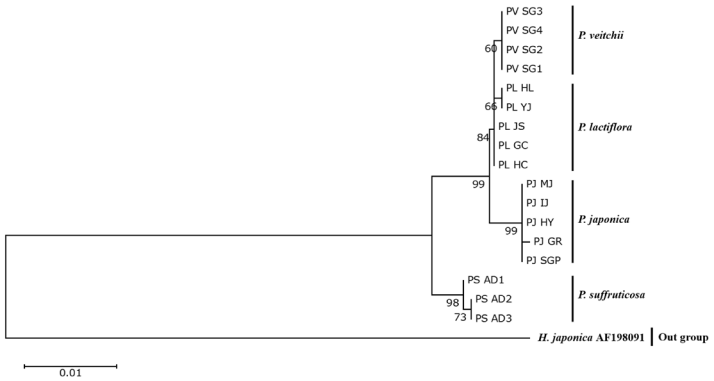

Figure 2. Phylogenetic tree with 1000 bootstrap replicates based on *mat*K sequences from four *Paeonia* species and *Hamamelis japonica*. NJ bootstrap values greater than 60% are shown above the branches.

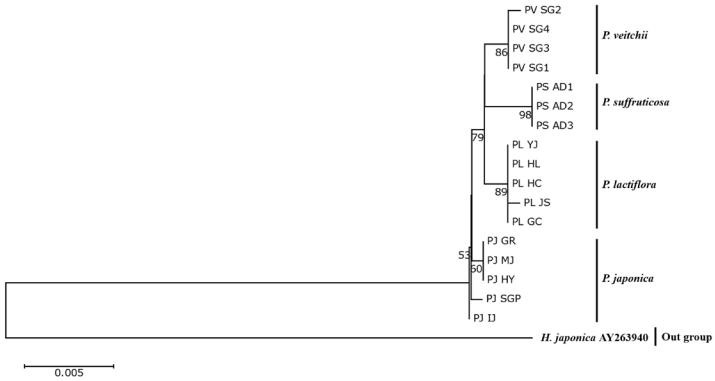

Figure 3. Phylogenetic tree with 1000 bootstrap replicates based on *rbc*L sequences from four *Paeonia* species and *Hamamelis japonica*. NJ bootstrap values greater than 50% are shown above the branches.

DISCUSSION

Diverse technical methods, including analyses of morphology, chemical composition, and molecular markers, have been used to identify medicinal plant species (Semagn et al., 2006; Jiang et al., 2014; Zhao et al., 2015). Due to limitations in the discriminatory power of morphological and chemical analysis, DNA barcoding has been suggested as a SNP marker for efficient and accurate species identification using short regions of nuclear or organellar DNA (Joshi et al., 2004; CBOL Plant Working Group, 2009; Jiang et al., 2014; Ma et al., 2014; Vassou et al., 2015). The ideal DNA barcode regions for this purpose should be applicable to diverse plant species. Therefore, the main focus of DNA barcoding is the discovery of universal sequence regions that contain conserved sequences, as well as diversity across species. Universal primers are designed to be complementary to the conserved sequences and flanking sites of the variable barcoding regions (Olmstead and Reeves, 1995; Kato et al., 1998; White et al., 1990; CBOL Plant Working Group, 2009). Furthermore, barcode sequences

should vary among species but not within species (Ma et al., 2014). In this study, we assessed rDNA-ITS, *mat*K, and *rbc*L as candidate DNA barcode regions for the identification of four *Paeonia* species.

The results of PCR amplification revealed that the three DNA barcode regions all had 100% amplification and sequencing efficiency, and are suitable for amplification in all four *Paeonia* species using universal primers. The inter- and intra-specific distances of rDNA-ITS were greater than those of *mat*K and *rbc*L (Table 2). Previous study has also shown that the inter-specific sequence divergence among DNA barcode regions is highest in ITS2, followed by *mat*K and *rbc*L, in *Sida cordifolia* and *Tulipa edulis* (Ma et al., 2014; Vassou et al., 2015). DNA barcode regions in plant systems are divided into protein-coding and non-coding regions. Non-coding regions, such as the nuclear ribosomal internal transcribed spacer (rDNA-ITS) and the *psb*A-*trn*H and *trn*L-F intergenic spacers, evolve quickly by acquisition of substitutions and indels, whereas protein-coding regions such as *mat*K and *rbc*L do not (Tripathi et al., 2013; Vassou et al., 2015). Our measured values of inter- and intra-specific distances are consistent with those reported previously and they confirm that the rDNA-ITS locus has evolved more rapidly than *mat*K and *rbc*L in the four *Paeonia* species (Tripathi et al., 2013; Aziz et al., 2015; Vassou et al., 2015; Zhao et al., 2015).

Based on substitutions and indels, we obtained marker nucleotides that can be used for discrimination at the species level. Such marker nucleotides were present in rDNA-ITS and *rbc*L, but not in *mat*K. Although *mat*K is one of the most rapidly evolving protein-coding regions in plastids (Tripathi et al., 2013; Ma et al., 2014; Vassou et al., 2015), *mat*K of *P. lactiflora* contains no marker nucleotides and consequently cannot be used to distinguish *P. lactiflora* from the other three *Paeonia* species. Although *mat*K has no discriminatory power, it could still be utilized as a DNA barcode region. Consequently, CBOL recommends the combination of *mat*K+*rbc*L as a universal DNA barcode for plants, with rDNA-ITS and *psb*A-*trn*H used as supplementary barcodes (CBOL Plant Working Group, 2009). Consistent with the CBOL recommendation, multiple studies reported that inter-specific discrimination requires a combination of two or more DNA barcode regions, e.g., ITS2+*psb*A-*trn*H or *rbc*L+ITS+*psb*A-*trn*H (Tripathi et al., 2013; Vassou et al., 2015). Therefore, although individual DNA barcode regions have sufficient discriminatory power, we nevertheless strongly recommend that combinations of DNA barcode regions, such as ITS+*rbc*L, ITS+*mat*K, or ITS+*mat*K +*rbc*L, can be used to achieve more accurate identification.

In phylogenetic analyses, the 17 samples formed a monophyletic group within each species, with bootstrap values greater than 50%. In particular, using the rDNA-ITS barcode, samples could be classified into the four species. *mat*K and *rbc*L also facilitated efficient separation of the four *Paeonia* species. In sequence analysis of *mat*K, which has no *P. lactiflora*-specific sequence, the phylogenetic tree exhibited better species identification resolution than the marker nucleotide analysis. In contrast, the phylogenetic tree of *rbc*L yielded poorer species resolution than did the marker nucleotide analysis in *P. japonica*. As an alternative method, we suggest using the *mat*K+*rbc*L combination as recommended by the CBOL plant working group.

Although DNA barcoding provides a useful method for discriminating the species, this technique takes 2-4 days and requires the following processes: PCR, cloning, sequence analysis, and alignment with GenBank. Therefore, further studies are needed to develop more rapid and efficient techniques for this species, such as melting curve analysis and multiplex-sequence characterized amplified region markers.

In conclusion, based on the assessment of all results including inter/intra-specific distance, marker nucleotides, phylogenetic trees, and amplicon sizes, the most suitable DNA barcode for identification of *Paeonia* species was rDNA-ITS. We also suggest that the *mat*K+*rbc*L combination also can be used to efficiently discriminate between the four *Paeonia* species. On the basis of these findings, we propose that the three DNA barcode regions, rDNA-ITS, *mat*K, and *rbc*L, can be used to discriminate between the four *Paeonia* species. Furthermore, species-specific marker nucleotides identified in this study will help to accurately authenticate each species and standardize the quality of the two traditional herbal medicines, Paeoniae Radix and Moutan Radicis Cortex.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We thank the 'Classification and Identification Committee of the KIOM' for critical authentication of plant materials and the Herbarium of Korea Standard Herbal Resources (herbarium code KIOM) for the provision of plant materials. Research supported by the grant 'Development of Basic Technologies for the Domestic Cultivation of Herbal Medicine Resources' (#K14417 and #K15421) from the Ministry of Science, ICT, and Future Planning (MSIP) of Korea to the KIOM.

REFERENCES

Aziz NAA, Ahmad MI and Naim DM (2015). Molecular DNA identification of medicinal plants used by traditional healers in Malaysia. *Genet. Mol. Res.* 14: 15937-15947. http://dx.doi.org/10.4238/2015.December.7.5

CBOL Plant Working Group (2009). A DNA barcode for land plants. *Proc. Natl. Acad. Sci. USA* 106: 12794-12797. http://dx.doi.org/10.1073/pnas.0905845106

Chen S, Pang X, Song J, Shi L, et al. (2014). A renaissance in herbal medicine identification: from morphology to DNA. *Biotechnol. Adv.* 32: 1237-1244. http://dx.doi.org/10.1016/j.biotechadv.2014.07.004

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41: 95-98.

Jiang C, Cao L, Yuan Y, Chen M, et al. (2014). Barcoding melting curve analysis for rapid, sensitive, and discriminating authentication of saffron (*Crocus sativus* L.) from its adulterants. *BioMed Res. Int.* 2014: 809037. http://dx.doi.org/10.1155/2014/809037

Joshi K, Chavan P, Warude D and Patwardhan B (2004). Molecular markers in herbal drug technology. *Curr. Sci.* 87: 159-165.

Kato H, Oginuma K, Gu Z, Hammel B, et al. (1998). Phylogenetic relationships of Betulaceae based on *mat*K sequences with particular reference on the positions of *Ostryopsis. Acta Phytotax. Geobot* 49: 89-97.

Kim J, Lee H, Lee Y, Oh BG, et al. (2007). Inhibition effects of Moutan Cortex Radicis on secretion of eotaxin in A549 human epithelial cells and eosinophil migration. *J. Ethnopharmacol.* 114: 186-193. http://dx.doi.org/10.1016/j.jep.2007.07.038

Kim MK, Wang H, Kim YJ, Sathiyamoorthy S, et al. (2013). Molecular authentication by multiplex-PCR of three similar medicinal plant species: *Cynanchum wilfordii, Cynanchum auriculatum* and *Polygonum multiflorum* (*Fallopia multiflorum*). *J. Med. Plants Res.* 7: 2584-2589.

Kimura M (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16: 111-120. http://dx.doi.org/10.1007/BF01731581

Korea Institute of Oriental Medicine. (2015). Defining Dictionary for Medicinal Herbs 'Hanyak Giwon Sajeon'. Available at [http://boncho.kiom.re.kr/codex/]. Accessed November 20, 2015.

Li X, Wang Y, Sun Q, Xu B, et al. (2016). Molecularly imprinted solid phase extraction in a syringe filter for determination of triazine herbicides in Radix Paeoniae Alba by ultra-fast liquid chromatography. *Talanta* 148: 539-547. http://dx.doi.org/10.1016/j.talanta.2015.11.027

Ma H, Zhu Z, Zhang X, Miao Y, et al. (2014). Species identification of the medicinal plant *Tulipa edulis* (Liliaceae) by DNA barcode marker. *Biochem. Syst. Ecol.* 55: 362-368. http://dx.doi.org/10.1016/j.bse.2014.03.038

Mahadani P and Ghosh SK (2013). DNA barcoding: A tool for species identification from herbal juices. *DNA Barcodes (Berlin)* 1: 35-38.

Moon BC, Ji Y, Lee YM, Kang YM, et al. (2015). Authentication of *Akebia quinata* DECNE. from its common adulterant medicinal plant species based on the RAPD-derived SCAR markers and multiplex-PCR. *Genes Genomics* 37: 23-32. http://dx.doi.org/10.1007/s13258-014-0225-6

Olmstead RG and Reeves PA (1995). Evidence for the polyphyly of the Scrophulariaceae based on choloroplast *rbc*L and *ndh*F sequences. *Ann. Mo. Bot. Gard.* 82: 176-193. http://dx.doi.org/10.2307/2399876

Semagn K, Bjørnstad Å and Ndjiondjop MN (2006). An overview of molecular marker methods for plants. *Afr. J. Biotechnol.* 5: 2540-2568.

Tamura K, Stecher G, Peterson D, Filipski A, et al. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30: 2725-2729. http://dx.doi.org/10.1093/molbev/mst197

Tripathi AM, Tyagi A, Kumar A, Singh A, et al. (2013). The internal transcribed spacer (ITS) region and trnH-psbA [corrected] are suitable candidate loci for DNA barcoding of tropical tree species of India. *PLoS One* 8: e57934. http://dx.doi.org/10.1371/journal.pone.0057934

Vassou SL, Kusuma G and Parani M (2015). DNA barcoding for species identification from dried and powdered plant parts: a case study with authentication of the raw drug market samples of *Sida cordifolia. Gene* 559: 86-93. http://dx.doi.org/10.1016/j.gene.2015.01.025

Wang Y, Wang P, Xu C, Sun S, et al. (2015). Discrimination and chemical characterization of different *Paeonia lactifloras* (Radix Paeoniae Alba and Radix Paeoniae Rubra) by infrared macro-fingerprint analysis-through-separation. *J. Mol. Struct.* 1099: 68-76. http://dx.doi.org/10.1016/j.molstruc.2015.06.035

White TJ, Bruns T, Lee S and Taylor J (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: PCR Protocols: A guide to methods and applications. (Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds.). Academic Press, New YorK, pp. 315-322.

Zhao Y, Li Y, Liu Y and Yang YF (2015). DNA barcoding for efficient identification of *Ixiolirion* species (Ixioliriaceae). *Genet. Mol. Res.* 14: 1903-1910. http://dx.doi.org/10.4238/2015.March.13.19

Zheng WH, Zhuo Y, Liang L, Ding WY, et al. (2015). Conservation and population genetic diversity of *Curcuma wenyujin* (Zingiberaceae), a multifunctional medicinal herb. *Genet. Mol. Res.* 14: 10422-10432. http://dx.doi.org/10.4238/2015.September.8.3

Supplementary material

Table S1. Genetic distances determined using rDNA-ITS sequences from 17 samples of four *Paeonia* species.

Table S2. Genetic distances determined using *mat*K sequences from 17 samples of four *Paeonia* species.

Table S3. Genetic distances determined using *rbc*L sequences from 17 samples of four *Paeonia* species.