J Plant Biotechnol (2016) 43:181−188
DOI:http://dx.doi.org/10.5010/JPB.2016.43.2.181

ISSN 1229-2818 (Print)
ISSN 2384-1397 (Online)

Research Article

Simple sequence repeat marker development from *Codonopsis lanceolata* and genetic relation analysis

Serim Kim · Ji Hee Jeong · Hee Chung · Ji Hyeon Kim · Jinsu Gil · Jemin Yoo · Yurry Um · Ok Tae Kim · Tae Dong Kim · Yong-Yul Kim · Dong Hoon Lee · Ho Bang Kim · Yi Lee

Received: 11 May 2016 / Revised: 20 June 2016 / Accepted: 20 June 2016
ⓒ Korean Society for Plant Biotechnology

Abstract In this study, we developed 15 novel polymorphic simple sequence repeat (SSR) markers by SSR-enriched genomic library construction from *Codonopsis lanceolata*. We obtained a total of 226 non-redundant contig sequences from the assembly process and designed primer sets. These markers were applied to 53 accessions representing the cultivated *C. lanceolata* in South Korea. Fifteen markers were sufficiently polymorphic, and were used to analyze the genetic relationships between the cultivated *C. lanceolata*. One hundred three alleles of the 15 SSR markers ranged from 3 to 19 alleles at each locus, with an average of 6.87. By cluster analysis, we detected clear genetic differences in most of the accessions, with genetic distance varying from 0.73 to 0.93. Phylogenic analysis indicated that the accessions that were collected from the same area were distributed evenly in the phylogenetic tree. These results indicate that there is no correlative genetic relationship between geographic areas. These markers will be useful in differentiating *C. lanceolata* genetic resources and in selecting suitable lines for a systemic breeding program.

Keywords *Codonopsis lanceolata*, Marker, Medicinal plant , Simple sequence repeats (SSR), Genetic relation

[†]S. Kim and J. H. Jeong contributed equally to this work.

S. Kim[†] · H. Chung · J. Gil · J. Yoo · Y. Lee (✉)
Department of Industrial Plant Science & Technology, Chungbuk National University, Cheongju 28644, Korea
e-mail: leeyi22@hanmail.net

J. H. Jeong[†] · T. D. Kim · Y.-Y. Kim
Seed & Seedling Management Division, Korea Forest Seed and Variety Center, Chungju 27495, Korea

J. H. Kim · H. B. Kim
Life Sciences Research Institute, Biomedic Co., Ltd., Bucheon 14548, Korea

Y. Um · O. T. Kim
National Institute of Horticultural and Herbal Science, Rural Development Administration, Chungbuk 27709, Korea

D. H. Lee
Department of Bio-systems Engineering, Chungbuk National University, Cheongju 28644, Korea

H. B. Kim
Faculty of Biotechnology, College of Applied Life Sciences, Jeju National University, Jeju 63243, Korea

Introduction

Codonopsis lanceolata, commonly called the bonnet bellflower, is a dicotyledonous perennial plant. It is included in the family Campanulaceae, which has 42 species predominantly distributed in East, Central and South Asia (He et al. 2015). It is highly valued as a traditional medicinal plant and a very popular vegetable in East Asia, especially, in the countries such as China, Japan and South Korea (Yoo et al. 1989).

Many phytochemical studies have reported that *C. lanceolata* roots contain saponins (Lee et al. 2002), phenylpropanoids (Ushijima et al. 2008), alkaloids and triterpenes (Jung et al. 2006), flavonoids (He et al. 2011), and more. In addition, many studies report that the chemical compounds of *C. lanceolata* show medicinal effects, influencing the immune system, cancer (tumor growth prevention) and gastrointestinal function (Sathiyamoorthy et al. 2011). Ichikawa et al (2009) have reported that there are seven kinds of saponins in *C. lanceolata*. Therefore, it is speculated that there is remarkable diversity in the composition and constituents of chemicals within cultivated *C. lanceolata*.

Several DNA markers have been successfully employed to analyze genetic relationships in *Codonopsis*. Doo et al (2002)

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

studied the genetic relationship of *C. lanceolata* collected from Baekdoo Mountain and Korea using random amplified polymorphic DNAs (RAPD). Lee et al (2001) also reported the discrimination and genetic relationship of *Adenophorae tryphylla* and *C. lanceolata* using RAPD method. Guo et al (2006) reported the successful development of inter-simple sequence repeats (ISSR) and RAPD methods and applied them to *C. lanceolata*. However, these methods were not sufficient to study genetic distance, and development of additional DNA markers is still needed.

Simple sequence repeat (SSR) marker is a powerful tool for analysis of genetic relationships. In addition, it is a useful tool for studying the non-reference plant genome, due to its even distribution throughout the genome, as well as its high polymorphism between individuals. Therefore, many studies have reported SSR marker applications to crops for phylogenetic analysis or genetic diversity on the non-reference plant genomes (Badiane et al. 2012; Bang et al. 2011; Kim et al. 2015; Park et al. 2013; Reed and Rinehart 2009). Li et al (2009, 2013) identified SSR markers for *C. tangshen* and *C. pilosula*, and then successfully applied these findings to investigate the genetic diversity and population structure of these two species. Although, *C. lanceolata* is one of the most important medicinal plants in Korea, an elite, inbred line or a variety has not been developed yet. Therefore, the study of genetic relationship or difference analysis should be carried out using markers based on genomic sequences.

In this study, we tried to develop novel SSR markers based on *C. lanceolata* genomic sequences to analyze the genetic relation of 53 cultivated accessions of *C. lanceolata,* collected from ten areas in South Korea.

Materials and Methods

Collection of Accessions and DNA Extraction

Fifty-three accessions of cultivated *C. lanceolata* fresh roots or seeds were collected from seed companies or farmers throughout South Korea. The collected accessions are listed in Appendix 1. All of the collected roots or seeds were grown in the Chungbuk National University greenhouse in the spring of 2015. For genomic DNA (gDNA) extraction, fresh leaves were ground with liquid nitrogen and kept in a deep freezer (-80°C) until gDNA extraction using the CTAB method (Doyle and Doyle 1987).

Development of SSR Markers

The microsatellite-enrichment library was constructed according to the method of Glenn and Schable (2005). Briefly, gDNA from *C. lanceolata* leaves (from accession CL0001) were digested with the restriction enzyme *Rsa* I to obtain DNA fragments ranging from approximately 300 to 1,000 base pairs (bp), and then ligated with a linker. The ligation products were subject to double enrichment steps by hybridization with 3'-biotinylated microsatellite probes. Information about 3'-biotinylated oligos for the enrichment was previously described by Glenn and Schable (2005). The DNA fragments, rich in microsatellite sequences, were ligated into the pGEM-T vector (Promega, Madison, WI, USA) and the ligation mixture was transformed into competent *E. coli* DH5α cells. The resulting colonies were subjected to colony PCR to identify recombinant clones using M13 forward and reverse primer sets. The PCR products were purified and used for sequencing by the ABI 3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA). After trimming the vector and linker sequences, nucleotide sequences were assembled to generate non-redundant contigs using Lasergene SeqMan (version 7.0.0, DNASTAR, Madison, WI, USA). Putative SSRs were identified by MISA software (http://www.pgrc.ipk-gatersleben.de/misa) using the following criteria: a minimum of three repeats for di-nucleotides to hexa-nucleotides and a gap within 100 bp for composite class. Criteria for primer design are as follows: 85-350 bp amplicon size and 57-60°C annealing temperature. Primers used in this study were synthesized by Biomedic Co., Ltd., Korea (www.ibiomedic.co.kr). The specificity of primers was validated by routine PCR using gDNA as the template. For the preliminary screening of polymorphic markers, routine genomic PCR was performed using gDNA from six selected accessions (CL0001, CL0004, CL0005, CL0006, CL0007, and CL0008) as templates. The PCR products were separated on a 2% agarose gel.

PCR Amplification and Genotyping

PCR was conducted using the Biometra Thermocycler (Göttingen, Germany) in a total volume of 20 ml containing 20 ng gDNA, 1 x HS™ Taq DNA polymerase buffer, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.2 mM of each primer, 1.25 units HS™ Taq DNA polymerase (Dongsheng Biotech, Guangzhou, China). The conditions for PCR amplification were as follows: 5 min for initial denaturation at 95°C, 34 cycles of 30 sec at 94°C, 30 sec at 57-61°C, 30 sec at 72°C, concluding with 1 cycle of 30 min at 72°C. PCR products were separated in a 2.0% agarose gel to visualize PCR amplification. PCR primer sets used in these experiments are listed in Table 1. Forward primers were labeled with a virtual dye 6-FAM, NED, VIC or PET (Applied Biosystems). After PCR amplification, 0.2 ml of PCR products were mixed with 9.8 ml Hi-Di formamide (Applied Biosystems) and 0.2 ml of GeneScan™ 500 LIZ® size standard

Table 1 Characteristics of the SSR markers developed and used in this study

No.	Marker ID	SSR motif	Annealing Temperature (°C)	Allele size range (bp)	Primer sequences (5'-3')	GenBank accession No.
1	CLSSR-1	(CATC)3(CATA)7	58	130-158	F : FAM-GACACAGCATTATCCACGAA R : GTTTAATTATGCGTTTTGGCTGTC	KP245963
2	CLSSR-2	(CT)8(ATCT)5	59	178-249	F : PET-CACCACTCAATCATGCAAGC R : GTTTGACGCAGRRGCAGAAAAGAA	KP246055
3	CLSSR-3	(GATT)6	59	225-249	F : PET-TGTCTATCCTTAATCTTGACAAGT R : GCTAGGATTTCGTCCACACC	KP246111
4	CLSSR-4	(CTGT)4	61	231-263	F : FAM-AGGTGGAAACGGTGTCTTTG R : GTTTGTCCACAGATGCCATTCGTA	KP246182
5	CLSSR-5	(TATG)5	58	182-202	F : FAM-GAGAACAATTATTTAAGAACGGATG R : GTTTGTCAGGCCCGRAAAAATGAA	KP246037
6	CLSSR-6	(AGAA)7	59	136-164	F : PET-TGGACTGTAGATGCCCTGCT R : GTTTTTAAGCCCAAGGTGTTCGTT	KP246090
7	CLSSR-7	(TATG)5	58	128-160	F : VIC-TAGTTTGGTGGGTAGGTGGA R : GTTTTGCACTTCCCAAGAAGAAAC	KP246032
8	CLSSR-8	(ATAC)6	60	107-143	F : NED-TTATGCAACTGCTAATTGGTTT R : GGAAAACTCCTGCTGCTGAA	KP246034
9	CLSSR-9	(CATA)7	58	161-173	F : VIC-GCAGAAGTGAGTATGCAAGTAG R : GTTTGATCATATCTATTGGCATGCA	KP246035
10	CLSSR-10	(TTGT)6	57	112-132	F : FAM-CATCCCTCCCTGAAAAATGT R : GTTTTGAAACTTTATGGGCATCTTG	KP246057
11	CLSSR-11	(AAGA)7	58	92-120	F : NED-GTCCTTGCTCACAATTAGCC R : GTTTTGACAAAATGGTGATGCCTA	KP246084
12	CLSSR-12	(AAGA)7	57	124-136	F : PET-GGGACGCATTTTCGTAATC R : GTTTAGAAATGTTTGTTTATGGGGTG	KP246086
13	CLSSR-13	(CTTTT)3C(TCTT)7	58	131-155	F : PET-AAGTGGTGAAATTGCAGAGG R : GTTTCATCAATGGATGCGAATAGA	KP245959
14	CLSSR-14	(TGAT)4T(GA)7	58	182-192	F : VIC-CCACTGGAACAAAGATTACGG R : GTTTCATGGAATTTTCATCGACAAGA	KP246040
15	CLSSR-15	(CAAA)4	59	207-235	F : NED-CCAATGTCTTGTGTTTGAGATG R : GTTTCAAGGCATTGTAACGTGCAT	KP246054

(Applied Biosystems). The mixture was denatured at 95°C for 5 min and placed on ice. The amplified fragments were separated by capillary electrophoresis on the ABI 3730 DNA analyzer (Applied Biosystems) using a 50-cm capillary with a DS-33 install standard as a matrix. We analyzed the amplicon size using the GeneMapper software (version 4.0, Applied Biosystems).

Data Analysis for Genetic Diversity

The informative bands were scored based on presence/absence/miss (1/0/9), and were used as the data set generation for analysis. The locus and variant of each SSR marker were analyzed. Genetic diversity (h) values of loci were calculated using the genetic diversity index. These values were calculated as follows: $h=1-\sum p_i^2$, where p_i is the frequency of the ith allele. In addition, heterozygosity of each locus was estimated for each SSR marker (Nei 1978). Polymorphism information content (PIC), a measure of closely related diversity, was estimated using PowerMarker software (version 3.25) (Botstein et al. 1980; Liu and Muse 2005). To analyze the genetic diversity of the collected accession, we performed statistical calculations using NTSYS software (version 2.11) (Rohlf 1992). The Jaccard genetic similarity matrix was used to

construct an Unweighted Pair Group Method with Arithmetic Average (UPGMA) dendrogram. WINBOOT software (Yap and Nelson 1996) was used for bootstrap analysis.

Results and Discussion

SSR Marker Development

The SSR enrichment approach has been widely used to isolate SSR markers efficiently from diverse organisms including plants (Glenn and Schable 2005; Zane et al. 2002). A library highly enriched for di-, tri-, and tetra-nucleotide types of SSR motifs was constructed from the genomic DNA of *C. lanceolata*. A total of 456 individual recombinant clones were subject to nucleotide sequencing by both directions using universal primers of the cloning vector. Finally, we obtained a total of 226 non-redundant contig sequences from the contig assembly process and deposited to the GenBank database under the accession numbers of KP245956 - KP246182. Primer sets were designed from the flanking sequences of the SSR motifs and used for the primary screening of polymorphic primer sets. To test polymorphism of the isolated SSR loci using the designed primer sets, total genomic DNA was isolated from genetic resources of various collection areas and was subjected to PCR amplification. Primary polymorphic SSR primers were selected based on agarose gel electrophoresis pattern of the PCR products. Finally, 15 polymorphic SSR primer sets based on *C. lanceolata* genomic sequences were identified (Table 1). The amplified band sizes ranged from 92 to 263 bp, and clearly showed single or double bands in electrophoresis. From these results, we obtained novel SSR markers based on genomic sequences to analyze *C. lanceolata* genetic resources.

Polymorphisms of the Developed SSR Markers

We obtained clearly amplified bands using 15 SSR markers from the 53 collected accessions. The polymorphisms of all samples were analyzed by GeneScan™ 500 LIZ® size standard. The 15 SSR loci identified were polymorphic. 103 unique alleles were detected from the 53 accessions, which varied from 3 to 19 alleles at each locus, with an average of 6.87 (Table 2). CLSSR-2 showed as many as 19 alleles. The genetic diversity for the loci tested in the total accessions showed a mean value of 0.62, varying from 0.14 (in CLSSR-5) to 0.86 (in CLSSR-2). The average of heterozygosity was 0.42, and CLSSR-2, CLSSR-3, CLSSR-8, CLSSR-11, CLSSR-15 showed > 0.5 (Table 2). PIC values were calculated for each polymorphic marker using a method that gives a maximum value of 0.50 (Roldán-Ruiz et al. 2000). The average PIC value was 0.57. CLSSR-2 exhibited the highest PIC value, 0.85. Five markers showed <0.5. The PIC value reflects the amount of polymorphism and is an informative marker if the value is >

Table 2 Genetic diversity measures for 15 polymorphic SSR loci in 53 *Codonopsis lanceolata* accessions

Marker ID	Number of alleles	Genetic diversity	Heterozygosity	PIC[a]
CLSSR-1	7	0.63	0.49	0.60
CLSSR-2	19	0.86	0.51	0.85
CLSSR-3	7	0.75	0.68	0.71
CLSSR-4	7	0.76	0.21	0.73
CLSSR-5	3	0.14	0.08	0.14
CLSSR-6	8	0.47	0.44	0.45
CLSSR-7	3	0.50	0.17	0.39
CLSSR-8	9	0.74	0.67	0.71
CLSSR-9	4	0.55	0.42	0.45
CLSSR-10	6	0.75	0.36	0.71
CLSSR-11	7	0.50	0.51	0.47
CLSSR-12	4	0.55	0.49	0.50
CLSSR-13	6	0.68	0.27	0.63
CLSSR-14	6	0.68	0.37	0.63
CLSSR-15	7	0.7	0.62	0.65
Mean	6.87	0.62	0.42	0.57

[a]PIC (Polymorphic Information Content) > 0.5 indicates an informative marker.

0.5. Although the PIC values of the five markers out of the 15 SSR markers were less than 0.5, they were sufficient to analyze the genetic diversity of the accession.

Genetic Relationship among Accessions

The genetic relationship among the accessions was analyzed, and an UPGMA cluster of the 53 *C. lanceolata* was constructed based on fragment analysis data using SSR markers. The genetic distance value ranged from 0.73 to 0.93, and there was no observed distinct group among the accessions (Fig. 1). Phylogenic analysis indicates that CL0009 through CL0020 accessions, which were collected in Hoengseong-gun, Gangwon-do, were distributed evenly in the phylogenetic tree. These results indicate that there is no correlative genetic relationship between the collection areas. CL0027 and CL0032 accessions, which were obtained from Yangpyeong-gun, Gyeonggi-do, and Yongin-si, Gyeonggi-do, respectively, were the only accessions having exactly the same genotype.

In this study, we developed 15 novel SSR markers based on *C. lanceolata* genomic sequences, and successfully applied them to the collected accessions, showing 103 polymorphic bands. Also, the PIC values indicated that the almost SSR markers were informative (Table 2). Whereas the developed

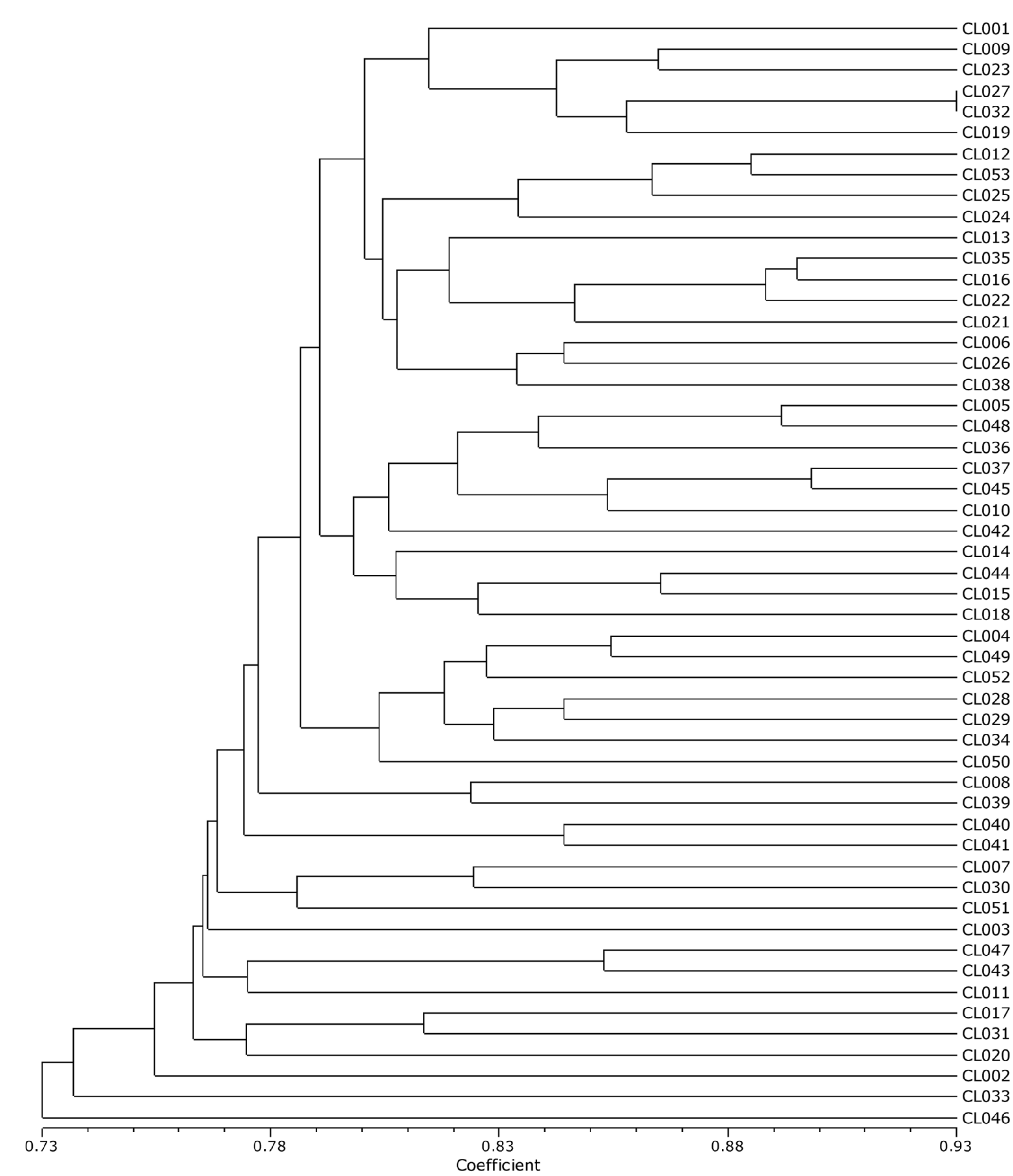

Fig. 1 Phylogenetic tree of the 53 *Codonopsis lanceolata* accessions created using UPGMA cluster analysis

SSR markers showed 103 polymorphic and reproducible bands, 20 RAPD primers generated only 49% of the polymorphic bands among the total PCR products (Doo et al. 2002). These results give evidence that the newly developed SSR markers are more efficient than RAPD. While the mean heterozygosity for the developed SSR markers was 0.42, Guo et al (2006) reported that the *C. lanceolata* heterozygosity for ISSR markers was extremely low. Therefore, we think that SSR markers are more useful than ISSR markers in heterozygous crops such as *C. lanceolata*. When the phylogenetic tree was constructed, we could not identify a distinct group or cluster to classify cultivated *C. lanceolata* (Fig. 1). These results indicate that the cultivated *C. lanceolata* plants had various genetic backgrounds and no variety had been developed yet.

Recently, the development of sequencing technology makes it possible to obtain high-throughput genetic information. Gao et al (2015) reported the transcriptome analysis of *C. pilosula* (Franch.) Nannf. using next generation sequencing (NGS) technology, and they provided the biosynthetic pathway of *Codonopsis* polysaccharides. This report suggested that it is possible to develop extensive genomic SSR, EST-SSR or single nucleotide polymorphism (SNP) markers from the crops have no reference genome. In the future, these markers would help to study the genomics or genetics of *C. lanceolata* cultivars as well as wild ones by facilitating genetic map construction, trait mapping, diversity studies, or selecting suitable lines for developing mapping populations and systemic breeding programs.

Acknowledgments

This work was carried out with the support of the "Research project of Korea Forest Seed and Variety Center (KFSV)" and the "Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ01102202)" by Rural Development Administration (RDA), Republic of Korea.

References

Badiane FA, Gowda BS, Cissé N, Diouf D, Sadio O, Timko MP (2012) Genetic relationship of cowpea (*Vigna unguiculata*) varieties from senegal based on SSR markers. Genet Mol Res 11:292-304

Bang KH, Jo IH, Chung JW, Kim YC, Lee JW, Seo AY, Park JH, Kim OT, Hyun DY, Kim DH, Cha SW (2011) Analysis of genetic polymorphism of Korean ginseng cultivars and foreign accessions using SSR markers. Korean J Med Crop Sci 19:347-353

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

Doo HS, Ryu JH, Lee KS, Li HL, Liu XH (2002) Analysis of genetic relationship by RAPD technique for *Codonopsis lanceolata* Trauty collected from the Baekdoo Mountain and Korea. Korean J Med Crop Sci 10:194-199

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11-15

Gao JP, Wang D, Cao LY, Sun HF (2015) Transcriptome sequencing of *Codonopsis pilosula* and identification of candidate genes involved in polysaccharide biosynthesis. PLoS ONE 10.2:e0117342

Glenn TC, Schable NA (2005) Isolating microsatellite DNA loci. Methods Enzymol 395:202-222

Guo WL, Gong L, Ding ZF, Li YD, Li FX, Zhao SP, Liu B (2006) Genomic instability in phenotypically normal regenerants of medicinal plant *Codonopsis lanceolata* Benth. et Hook. f., as revealed by ISSR and RAPD markers. Plant Cell Rep 25:896-906

He JY, Ma N, Zhu S, Komatsu K, Li ZY, Fu WM (2015) The genus *Codonopsis* (Campanulaceae): a review of phytochemistry, bioactivity and quality control. J Nat Med 69:1-21

He X, Zou Y, Yoon WB, Park SJ, Park DS, Ahn J (2011) Effects of probiotic fermentation on the enhancement of biological and pharmacological activities of *Codonopsis lanceolata* extracted by high pressure treatment. J Biosci Bioeng 112:188-193

Ichikawa M, Ohta S, Komoto N, Ushijima M, Kodera Y, Hayama M, Shirota O, Sekita S, Kuroyanagi M (2009) Simultaneous determination of seven saponins in the roots of *Codonopsis lanceolata* by liquid chromatography–mass spectrometry. J Nat Med 63:52-57

Jung SW, Han AJ, Hong HJ, Choung MG, Kim KS, Park SH (2006) α-glucosidase inhibitors from the roots of *Codonopsis lanceolata* Trautv. Agric Chem Biotechnol 49:162-164

Kim HJ, Yeo SS, Han DY, Park YH (2015) Interspecific transferability of watermelon EST-SSRs assessed by genetic relationship analysis of cucurbitaceous crops. Kor J Hort Sci Technol 33:93-105

Lee KT, Choi J, Jung WT, Nam JH, Jung HJ, Park HJ (2002) Structure of a new echinocystic acid bisdesmoside isolated from *Codonopsis lanceolata* roots and the cytotoxic activity of prosapogenins. J Agric Food Chem 50:4190-4193

Lee MY, Mo SY, Kim DW, Oh SE, Ko BS (2001) Discrimination and genetic relationship of *Adenophorae triphylla* (Thunb) A. DC. var. japonica Hara and *Codonopsis lanceolata* Trauty using RAPD analysis. Korean J Med Crop Sci 9:205-210

Li Z, Liu X, Wang X, Fan B, Wang X, Zhao G (2013) Isolation and characterization of novel microsatellite markers in *Codonopsis tangshen* (Campanulaceae). Conserv Genet Res 5:393-395

Li ZH, Wen HY, Chen J, Wu GL, Wang YJ (2009) Development of 10 polymorphic microsatellite loci primers for *Codonopsis pilosula* Nannf. (Campanulaceae). Conserv Genet 10:747-749

Liu KJ, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128-2129

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583-590

Park SH, Choi SR, Lee JS, Nguyen VD, Kim SG, Lim YP (2013) Analysis of the genetic diversity of radish germplasm through SSR markers derived from chinese cabbage. Korean J Hort Sci 31:457-466

Reed SM, Rinehart TA (2009) Simple-sequence repeat marker analysis of genetic relationships within *Hydrangea paniculata*. HortSci 44:27-31

Rohlf FJ (1992) NTSYS-pc:numerical taxonomy and multivariate analysis system. Applied Biostatistics

Roldán-Ruiz I, Dendauw J, Van Bockstaele E, Depicker A, De Loose M (2000) AFLP markers reveal high polymorphic rates in ryegrasses (*Lolium* spp.). Mol Breed 6:125-134

Sathiyamoorthy S, In JG, Lee OR, Lee BS, Devi SR, Yang DC (2011) *In silico* gene expression analysis in *Codonopsis lanceolata* root. Mol Biol Rep 38:3541-3549

Ushijima M, Komoto N, Sugizono Y, Mizuno I, Sumihiro M, Ichikawa M, Hayama M, Kawahara N, Nakane T, Shirota O, Sekita S, Kuroyanagi M (2008) Triterpene glycosides from the roots of *Codonopsis lanceolata*. Chem Pharm Bull 56: 308-314

Yap IV, Nelson RJ (1996) Winboot: a program for performing bootstrap analysis of binary data to determine the confidence limits of UPGMA-based dendrograms. International Rice Research Institute. Manila, Philippines. pp.1-22.

Yoo KO, Lee WT (1989) A taxonomic study of the genus *Codonopsis* in Korea. Kor J Plant Tax 19:81-102

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. Mol Ecol 11:1-6

Appendix Table 1 Accession names, collection types and collected regions of 53 *Codonopsis lanceolata* from South Korea for the genetic relationship analysis

No.	Accession name	Collection type	Collected region
1	CL0001	Root	Samcheok-si, Gangwon-do
2	CL0002	Seed	Wonju-si, Gangwon-do
3	CL0003	Seed	Cheolwon-gun, Gangwon-do
4	CL0004	Root	Chuncheon-si, Gangwon-do
5	CL0005	Root	Pyeongchang-gun, Gangwon-do
6	CL0006	Root	Backduck Mountain, Pyeongchang-gun, Gangwon-do
7	CL0007	Root	Maebong Mountain, Hongcheon-gun, Gangwon-do
8	CL0008	Root	Hongcheon-gun, Gangwon-do
9	CL0009	Root	Hoengseong-gun, Gangwon-do (1)
10	CL0010	Root	Hoengseong-gun, Gangwon-do (2)
11	CL0011	Root	Hoengseong-gun, Gangwon-do (3)
12	CL0012	Root	Hoengseong-gun, Gangwon-do (4)
13	CL0013	Root	Hoengseong-gun, Gangwon-do (5)
14	CL0014	Root	Hoengseong-gun, Gangwon-do (6)
15	CL0015	Seed	Hoengseong-gun, Gangwon-do (7)
16	CL0016	Seed	Hoengseong-gun, Gangwon-do (8)
17	CL0017	Seed	Hoengseong-gun, Gangwon-do (9)
18	CL0018	Seed	Hoengseong-gun, Gangwon-do (10)
19	CL0019	Seed	Hoengseong-gun, Gangwon-do (11)
20	CL0020	Seed	Hoengseong-gun, Gangwon-do (12)
21	CL0021	Seed	Hoengseong-gun, Gangwon-do (13)
22	CL0022	Seed	Hoengseong-gun, Gangwon-do (14)
23	CL0023	Seed	Hoengseong-gun, Gangwon-do (15)
24	CL0024	Seed	Hoengseong-gun, Gangwon-do (16)
25	CL0025	Seed	Namyangju-si, Gyeonggi-do
26	CL0026	Seed	Bucheon-si, Gyeonggi-do
27	CL0027	Root	Yangpyeong-gun, Gyeonggi-do
28	CL0028	Root	Yongin-si, Gyeonggi-do (1)
29	CL0029	Seed	Yongin-si, Gyeonggi-do (2)
30	CL0030	Seed	Yongin-si, Gyeonggi-do (3)
31	CL0031	Seed	Yongin-si, Gyeonggi-do (4)
32	CL0032	Seed	Yongin-si, Gyeonggi-do (5)
33	CL0033	Root	Jiri Mountain, Sancheong-gun, Gyeongsangnam-do (1)
34	CL0034	Seed	Jiri Mountain, Sancheong-gun, Gyeongsangnam-do (2)
35	CL0035	Seed	Bonghwa-gun, Gyeongsangbuk-do
36	CL0036	Root	Ulleung-gun, Gyeongsangbuk-do
37	CL0037	Root	Yeongju-si, Gyeongsangbuk-do
38	CL0038	Seed	Jung-gu, Daejeon Metropolitan City (1)
39	CL0039	Seed	Jung-gu, Daejeon Metropolitan City (2)
40	CL0040	Seed	Songpa-gu, Seoul
41	CL0041	Seed	Jongno-gu, Seoul
42	CL0042	Root	Gwangyang-si, Jeollanam-do
43	CL0043	Root	Jiri Mountain, Gurye-gun, Jeollanam-do
44	CL0044	Root	Sinan-gun, Jeollanam-do
45	CL0045	Root	Hwasun-gun, Jeollanam-do
46	CL0046	Root	Muju-gun, Jeollabuk-do (1)
47	CL0047	Root	Muju-gun, Jeollabuk-do (2)
48	CL0048	Root	Seogwipo-si, Jeju Special Self-Governing Province
49	CL0049	Root	Jeju-si, Jeju Special Self-Governing Province (1)
50	CL0050	Root	Jeju-si, Jeju Special Self-Governing Province (2)
51	CL0051	Seed	Jeju-si, Jeju Special Self-Governing Province (3)
52	CL0052	Root	Danyang-gun, Chungcheongbuk-do
53	CL0053	Root	Chungju-si, Chungcheongbuk-do