A THESIS FOR THE DEGREE OF MASTER OF SCIENCE

Development of Molecular Markers Based on AFLP and MITE in Ginseng Species

BY

SOUNGJUN LIM

AUGUST, 2012

MAJOR IN CROP SCIENCE AND BIOTECHNOLOGY

DEPARTMENT OF PLANT SCIENCE

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

Development of Molecular Markers

Based on AFLP and MITE

in Ginseng Species

UNDER THE DIRECTION OF DR. TAE-JIN YANG

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF SEOUL NATIONAL UNIVERSITY

BY

SOUNGJUN LIM

MAJOR IN CROP SCIENCE AND BIOTECHNOLOGY

DEPARTMENT OF PLANT SCIENCE

AUGUST, 2012

APPROVED AS A QUALIFIED DISSERTATION OF SOUNGJUN LIM

FOR THE DEGREE OF MASTER OF SCIENCE

BY THE COMMITTEE MEMEBERS

AUGUST, 2012

CHAIRMAN _____

SUK-HA LEE, Ph.D.

VICE-CHAIRMAN _____

TAE-JIN YANG, Ph.D.

MEMBER _____

DO-SOON KIM, Ph.D.

Development of Molecular Markers Based on AFLP and MITE in Ginseng Species

SOUNGJUN LIM

DEPARTMENT OF PLANT SCIENCE

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

ABSTRACT

Korean ginseng (*Panax ginseng* C. A. Meyer) is an important medicinal herb belonging to family *Araliaceae*. Three species, *P. ginseng, P. notoginseng* and *P. quinquefolius,* have been generally used as herbal medicine worldwide. To date, nine *Panax ginseng* cultivars have been cultivated. However, identification of them is being carried out only by empirical morphological observation. Thus, the main

objective of this study is development of AFLP and MITE-based molecular markers to distinguish *P. ginseng* cultivars and related species. AFLP analysis is an excellent technique for molecular marker development of plant species whose genome information is limited. In AFLP analysis, 117 (21.3%) polymorphic bands were identified between *P. ginseng* and *P. quinquefolium* and 5 (0.9%) bands among *P. ginseng* cultivars among the total 549 amplified bands. This inefficiency result to find a marker among ginseng cultivars may be due to low genetic variation among cultivars. MITE which has been successfully used in other plant species can also be used for marker development in ginseng species, because of its characteristic small size and high copy number. In this study, 133 MITE consensus sequences were identified. As a result, 73 MITE regions of 25 MITE families were analyzed and 16 (21.9%) polymorphic regions were identified between *P. ginseng* and *P. quinquefolium*. In addition, 10 MITE regions were identified to show MITE-based insertional polymorphism (MIP) patter, 3 between *P. ginseng* and *P. quinquefolium* and rest 7 within single cultivar itself. All MIP bands found in *P. quinquefolium* were smaller than their counterpart bands in *P. ginseng*. This implies that MITEs might be inserted in *P. ginseng* genome after divergence from *P. quinquefolium*. Furthermore, many InDel and base substitutions were found among flanking sequences of MITE, when compared scaffold sequences matching to MIP amplicons. It indicates that *P. ginseng* genome was probably duplicated before insertion of the MITEs, although some MIP regions found in scaffolds seem to be resulted from MITE excision. In conclusion, this study analyzed genetic polymorphism in *P. ginseng* cultivars and related *Panax* species by AFLP and MITE analysis, and revealed polymorphic regions that can be used for molecular marker development. These results will be a valuable resource to understand structure and evolution of ginseng genome.

Key words: *P. ginseng, P. quinquefolium, Araliacea*, AFLP, MITE, MIP, marker

Student number: 2010-23401

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AFLP : Amplified Fragment Length Polymorphism

CAPS : Cleaved Amplified Polymorphic Sequences

InDel : Insertion or Deletion

MIP : MITE Insertion Polymorphism

MITE : Miniature Inverted-repeat Transposable Element

PCR : Polymerase Chain Reaction

SNP : Single Nucleotide Polymorphism

SSR : Simple Sequence Repeat

TE : Transposable Element

TIR : Terminal Inverted Repeat

TSD : Target Site Duplication

INTRODUCTION

Korean ginseng (*Panax ginseng* C. A. Meyer) is one of the most important medicinal plants belonging to the relict family *Araliaceae*. Its roots have been used as a popular oriental medicine for thousands of years in Eastern Asia including Korea and China [1]. Among 15 species in the genus *Panax,* three species, *P. ginseng, P. notoginseng* and *P. quinquefolius,* have been widely cultivated commercially because of their unique pharmacological effects such as immune system stimulant, anti-carcinogenic activity and beneficial to the central nervous system [2, 3].

So far, nine cultivars, ‘Chunpoong’, ‘Chunsun’, ‘Yunpoong’, ‘Gopoong’, ‘Sunpoong’, ‘Sunun’, ‘Sunwon’, ‘Sunhyang’ and ‘Gumpoong’, have been cultivated by pure line selection and registered as commercial varieties. However, identification among these cultivars is being carried out only by empirical morphological observation. Thus scientific approaches such as molecular markers have always been required to identify ginseng cultivars and to develop new cultivar with excellent agricultural and pharmacological traits.

Until now, several kinds of markers have been tried to identify *P. ginseng* cultivars: SNP markers [4, 5], CAPS marker [6], and SSR marker [7]. Nevertheless, none of molecular marker has been developed to distinguish *P. ginseng* cultivars completely.

Meanwhile, several studies have been reported for study on polymorphism of DNA in *P. ginseng*. For examples, amplified fragment length polymorphism (AFLP) analysis was employed to find DNA polymorphism and to develop marker in *P. ginseng* [8, 9]. In addition, transposable elements (TEs) including Miniature inverted-repeat transposable elements (MITEs) are being analysed using *P. ginseng* genome sequence by a part of Korea Ginseng Genome Sequencing Project (http://im-crop.snu.ac.kr/new/index.php).

AFLP is a molecular marker technique based on selective amplification of the DNA restriction products [10]. AFLP has been widely used to investigate DNA polymorphism and to develop molecular markers for plant species whose genome sequence was not available, because this technique does not require genome sequence information.

MITEs were first discovered in maize [11] and are a special type of Class 2 non-autonomous TEs. MITEs are characterized by their relatively small size (< 600 bp), terminal inverted repeats (TIRs), flanking target site duplications (TSDs) and their relatively high copy numbers compared to other types of Class 2 TEs [12]. MITEs transpose through a 'cut-and-paste' mechanism, but they don't have coding internal sequences for transposition, unlike autonomous TEs. Instead, MITEs might get their transposition activity by using transposases encoded by other autonomous DNA transposons [13, 14]. MITEs have been identified in many eukaryotic genomes including sorghum [15], rice [16] and human [17]. In plants, most MITEs are classified as *Tourist*-like element with a 3bp TSDs (usually TTA

or TAA) or *Stowaway*-like element with a 2bp TSDs (usually TA) [18-20]. Several MITEs could form a hairpin loop secondary structures because of TIRs that extend to almost half the length of the MITEs [21]. To date, several programs have been developed to identify MITE sequence: TRANSPO [22], FINDMITE [23], MUST [24] and MITE-Hunter [25]. Although these programs have their-own merits and demerits to find specific MITE sequence, MITE-Hunter is generally considered to be the best discovery tool in *de novo* identification of MITEs [26].

The main objective of the present study is to develop AFLP- and MITE-based molecular markers to distinguish *P. ginseng* cultivars and related species. For this, I firstly analysed AFLP and found polymorphic DNA fragments. Next, I characterized MITE sequences in *P. ginseng* draft whole genome sequence and found polymorphic DNA region with MITE. In addition to the objective, I elucidated a reason why MITE-based insertional polymorphism occurred by comparison among *P. ginseng* scaffold sequences. To my best knowledge, the present study is the first report of AFLP- and MITE-based polymorphism analysis for nine *P. ginseng* cultivars and related *Panax* species.

MATERIAL AND METHODES

1. Plant materials and genomic DNA extraction

Eleven *P. ginseng* plant samples were used to extract genomic DNAs: nine registered cultivars, 'Chunpoong', 'Chunsun', 'Yunpoong', 'Gopoong', 'Sunpoong', 'Sunun', 'Sunwon', 'Sunhyang' and 'Gumpoong', cultivated by inbred line selection in Korea Ginseng Corporation (KGC) Natural Resources Research Institute (Daejeon, Korea), one collected from Russia, and one Japanese cultivar, 'Mimaki' collected from Japan. In addition to these, American ginseng (*P. quinquefolium*) originated in the USA, was also used. All leaf samples were kindly provided from KGC Central Research Institute. Total Genomic DNA was extracted using the modified cetyltrimethylammonium bromide (CTAB) method [27]. DNA concentration was measured by ND-1000 (Nanodrop Technologies Inc., USA) and adjusted to 100 ng/ml. The DNA samples were stored at -20℃ before using.

2. AFLP analysis

AFLP analysis was performed as described in [28]. The EcoR1 and Mse1 restriction endonucleases were used for digesting two microgram of genomic DNA from each sample and then adaptor sequences were ligated at both end of digested DNA fragments. And the adaptor-ligated DNA was ten

times diluted with sterilized water, and 2.5 μL of them were used as template for pre-amplification. Pre-amplifications were performed in a total volume of 25 μL containing 2.5 μL of diluted ligated solution, 5 mM of each dNTP, 10 pmol of EcoR1 +1 primer and Mse1 +1 primer, and 2 units *Taq* DNA polymerase (Vivagen, Korea) using a DNA Engine Thermal Cycler (Bio-rad, USA). The polymerase chain reaction (PCR) cycling conditions were as follows: 2 min at 95℃ for denaturation, 30 cycles of 30 sec at 95℃, 30 sec at 56°C, 1 min at 72℃, and 5 min at 72℃ for final extension. The pre-amplified DNA was diluted 1:8 and then used for selective amplification.

Selective amplifications were performed in a total volume of 20 μL containing 2 μL of diluted pre-amplified DNA, 5 mM of each dNTP, 10 pmol of EcoR1 +3 primer and Mse1 +3 primer, and 1 units *Taq* DNA polymerase (Vivagen, Korea) using a DNA Engine Thermal Cycler (Bio-rad, USA). The PCR cycling conditions were as follows: 2 min at 95℃ for denaturation, 8 cycles of 30 sec at 95℃, 30 sec at 65°C that was reduced by 1℃ per cycle, 1 min at 72℃, 24 cycles of 30 sec at 95℃, 30 sec at 56°C, 1 min at 72℃, and 5 min at 72℃ for final extension.

The selective PCR product was separated by electrophoresis in 5% denaturing polyacrylamide gel at 90W (maximum of 2400V) for 2 h in T-Rex ™ Aluminum Baked Sequencer (model S3S, Owl separation systems, USA) in 1 x Tris-borate EDTA (TBE) buffer, followed by silver staining

using a Promega staining Kit (Promega, USA). The size of the scored PCR product was in range between 50 and 2,200 bp.

3. Identification of MITE sequence in *P. ginseng* scaffold

The *P. ginseng* genome sequence database containing 2.2 Gb of 'Chunpoong' cultivar genomic DNA sequence was constructed by Korea Ginseng Genome Sequencing Project (http://im-crop.snu.ac.kr/new/index.php) and used for this study (Table 1). MITE consensus sequences in *P. ginseng* scaffold sequence database were identified using the program MITE-Hunter [25]. To identify TIRs from MITE consensus sequences found by MITE-hunter, the program Einverted included in EMBOSS sequence analysis package [29] were used. Hypothetical secondary structures of MITE consensus sequence were predicted by MFOLD software [30, 31].

By Blastn search with MITE consensus sequences as queries, scaffold sequences containing MITE sequences were found and retrieved in *P. ginseng* genome sequence database. To amplify MITE containing sequence, polymerase chain reaction (PCR) primers were designed within approximately 150 bp from either TIR of intact MITE present, using the Primer3 software (http://frodo.wi.mit.edu/primer3/) (Table 2).

The putative copy number of MITE consensus sequence was determined based on BlastN hit number against *P. ginseng* genome sequence.

Table 1. Ginseng genome assembly (Basic assembly using SOAP de novo/ NGS Cell/ Allpaths-LG)

Assembly Information

Program: **NGS cell 4.06** Note: 1st,2nd,3rd,4th,contig and MP

	Contigs No.	Contigs Length	N50	Largest	Average	N's No.
All	2,546,489	3,277,430,424	3,638	129,722	1,287	942,827,307
0.5kbp over	1,180,908	2,817,520,631	5,323	129,722	2,386	893,933,125
1kbp over	569,933	2,393,614,912	8,212	129,722	4,200	838,715,968
2kbp over	281,351	1,995,782,020	11,075	129,722	7,094	807,499,168
Scaffolding	229,726	2,209,607,684	32,750	322,333	9,618	398,126,132

Table 2. Primers used for MITE insertional polymorphism search of *P. ginseng* scaffolds

MITE family	Members	Primer	Forward (5'-3')	Reverse (5'-3')
pg2	pg2-1	pgM1-1	ATATTGGGGCCATAAACCAT	CACAAAGTCCCAGCCGTAAT
	pg2-2	pgM1-2	TTTTTGCGTGTGGGAACTAA	GCTTATTTGGTCCGTTGGAG
	pg2-3	pgM1-3	TGTTCAAAACTCAGAAACAGAGTG	TGTGGATCAATCAGTACAACCTG
	pg2-4	pgM1-4	GGTTGGATGGAATGGACCTA	GGAAGCATTTAACCCTAATTCAGA
	pg2-5	pgm02-1(1)	TGTCAATCAAAAACAATCTACACAAA	ATGCCTCGCCAGGATAAGT
	pg2-6	pgm02-2(2)	CCCCTTTATGGGAAGTTGAA	GAGATGGGGTCCATTAATTTTT
pg3	pg3-1	pgm3-1-1	AAGAGATTTCTTTAGTCAATGTGTCAA	CATAACAGTGGTGGATTCAGGA
pg4	pg4-1	pgm4-1-2	TTTTACCAACTCTAGTTTTTCTTTTTC	TTATTTTTGATAGGTCGACGTTGT
	pg4-2	pgm4-2-3	CACCATGTCAGAGGTAAATGTG	TCGGATCCTAAGAGTTTTGTCA
	pg4-3	pgm4-3-4	GAAAGAAATATGAATGAAATTGAGG	CGGTTTCCACTGCTACGAA
pg6	pg6-1	pgm6-1-5	TGTGCTGTTCAACCCTTCTT	GAAGGCCCTTTCTCACAAAT
pg7	pg7-1	pgM2-1	TGCTCCATTGTCACCCATTA	ACGAGCTCACCCAAAAAGTC
	pg7-2	pgM2-2	TTCTGAAGGCAGCAATTTCA	ATGCTCCTTTGTCACCCATT
	pg7-3	pgM2-3	AGCGGACTTAGATCCCCTTC	TCACCCACTTTATCCCCAAC
	pg7-4	pgM2-4	GGTTCTAGATAGTGGATCCTAGCC	CCCCAAGGATTTCATTAGTTTT
pg12	pg12-1	pgm12-1-6	ATCAGCGTAGGAGGCTTTCA	AAGAGAAAACGGCACAAGGA
	pg12-2	pgm12-2-7	AGCGCATGAAGATCAGTTTG	CACTCTATATTGTCCGTCGCTAA
pg14	pg14-1	pgM3-1	CAACTGAGCAGCGACAGAAG	GTTGCAAGATGGATGTCCTG
	pg14-2	pgM3-2	CCTCAACGTTTCCAATGACA	TTTCTTCCATATTGTCCTACATCAA

Table 2. (Continued)

MITE family	Members	Primer	Forward (5'-3')	Reverse (5'-3')
pg14	pg14-3	pgM3-3	TGATCGCAGCATCCAAAA	GGAAGGCTTCAATTGCATAA
	pg14-4	pgM3-4	TGATCGCAGCATCCAAAA	GGAAGGCTTCAATTGCATAA
	pg14-5	pgm14-1-8	ATCTATGACTACTAAACGGGTGTGG	AGGACCGTAAAGCATTTTCC
	pg14-6	pgm14-2-9	AGAAGCAGCAAGATGCATGA	GAAGGATTATTTGATGGTATCGAC
	pg14-7	pgm14-3-10	GGCCTCCATAGGGAAAAATC	CTCTTCGATGCCACGTCATA
	pg14-8	pgm14-4-11	TTCCTTTTGCTAATGTATTTTGTG	TCAAATTCCGCTAAATATGCAA
pg16	pg16-1	pgM4-1	TTCATCGACTGACATGCTAGG	CAAGGGTGGGTGCTAAGTTT
	pg16-2	pgM4-2	ATCGGATGGGTACAGGTTAAT	TCAAAGCCATTTCCAATCCT
	pg16-3	pgM4-3	AAGTGTTCTTTATCTCTTTACATAGGG	TGCACGACAGTTGTATTAATGAAG
	pg16-4	pgM4-4	CAATTTTAAATATCTCAAACCATTGC	TTGTAATATTTTTCAATGAGGGATAA
pg17	pg17-1	pgM5-2	TGGGTGATGCACATGTTTTA	GCATTGCAAGCATTCTAACC
	pg17-2	pgM5-3	TGAGAATTGAGAAACTGTTGAGTTG	CCTTCAAGACCTGAGCCATT
	pg17-3	pgM5-4	ACACTTTCGGGTCACAGGTT	TGTAAAACAAACAAGTTCAGTTCAA
	pg17-4	pgm17-1-12	GGCACTGCGATATGTAAGCA	CAATATTATTTCAATATGCCAAGAAA
	pg17-5	pgm17-2-13	TCAGAATTTACCTGATACCTAACAGTC	TTATTTCTAGAATTGTCAACTTGTTTG
	pg17-6	pgm17-3-14	CCAAACCCCCGTATATCAAA	GGGTTGGTGGGAGTTAAACA
	pg17-7	pgm17-4-15	TCAGCAGATGCTCATCACAA	TTGCTTAACAAACCAAATCCTTC
pg20	pg20-1	pgm20-1-16	ACACACCCACCCTGATTTTC	TCGTGGGATTAATATTTTGTGA

Table 2. (Continued)

MITE family	Members	Primer	Forward (5'-3')	Reverse (5'-3')
pg24	pg24-1	pgm24-1-17	AATGCCTCTTGATTTCGACAC	GGTAAAATTTCATCCCATTCCA
	pg24-2	pgm24-2-18	CTTCCCATCAACCTTGTGTTC	AAGAAAGTCAGTATGATGACTACCAAA
	pg24-3	pgm24-3-19	TTAGGAACTGCAAAACCAAAA	TGGGTGAACTCTGAGCCTTT
	pg24-4	pgm24-4-20	CATATTCTGAGAAAAAGTGAAAGCA	CCCTTTCACATCCACCATTT
pg26	pg26-1	pgM6-1	CTCTGTGCATTGTGTGATTGAA	GGCGATTGACTAGGAAACTGA
	pg26-2	pgM6-2	AGGGATGTCAAGAAATTCACAGA	GATTCTAACAGAAATTTCCAAGCA
	pg26-3	pgM6-3	ATGCGGTGTTCCTCATTTTT	CCGAACCGGTACATATTAAATCA
	pg26-4	pgM6-4	TCACACCGCCTTGTAATTTTT	GAAAACAGGTGGGACCATTC
pg34	pg34-1	pgm34-1(3)	TTGAGCGATTCTACTTGTTCTACTG	TTGGGATTAGGTCACATATCCTT
	pg34-2	pgm34-2(4)	CAGCAACTCACTCCAAAGACA	TCGGTAAGGTGATTTTAGGAAGA
pg36	pg36-1	pgm36-1(5)	GAAATTGTGGACCAGGATAGC	GAAAACTTTTGGGTGATGGTT
pg37	pg37-1	pgm37-1-21	CTAGTCCGCACTTGAACAGC	AGCTCCAGCGAGATTTTCAG
pg42	pg42-1	pgm42-1-22	AAAAACAAATTTCCAGAAATCAATA	GTCGCTAAAGGGGGATGAG
	pg42-2	pgm42-2-23	AAAAGCATGTGTGAATTTAATTTTT	GAAACCAATTCTCATATTCATGTCA
pg46	pg46-1	pgM7-1	TCCAAGACACAAAATCAACACA	CCTGAAATCACAAACGCTGT
	pg46-2	pgM7-2	TCCAAGACACAAAATCAACACA	CCTGAAATCACAAACGCTGT
	pg46-3	pgM7-3	GTTCGTGTGCATGCTCCTT	GGTGGGCTTGTCTTAAGTGG
	pg46-4	pgM7-4	GGTTCTTTTCTTCTCGCAGGT	ATGCAAGCATGGTATGATGG

Table 2. (Continued)

MITE family	Members	Primer	Forward (5'-3')	Reverse (5'-3')
pg47	pg47-1	pgM8-1	CGTGGATTCTGTGATTGGAC	GCAGTCTGGCTGTGTAGCTG
	pg47-2	pgM8-2	CAATCGAACATGTCAAAGCA	CGAACCTTCAGGTAAAACAACC
	pg47-3	pgM8-3	CAGTCGAACATGTCAAAGCA	CGAACCTTCAGGTAAAACAACC
	pg47-4	pgM8-4	ACCAAATTAGGATGCCCAAC	ACGGGGCTCATTAGAAGATT
	pg47-5	pgM8-5	TTGCCCCACTTTGATGATAA	GCGGGGCTCATTAGAAGAT
pg48	pg48-1	pgm48-1(6)	TGTCGTAGTCTGCAAGGTTTTT	TCCAAGGACATCTATAACCCAAA
pg51	pg51-1	pgm51-1-24	CGTAGATGCATCACCAGTTTC	TCCAATTTTACCGTTGAGTGG
	pg51-2	pgm51-2-25	CAAGGAACCCCATCCACTAT	GGTTTCTGGTTCAATTGGAATATC
pg61	pg61-1	pgM9-1	GAAGACCAAAGCGGAGAAGA	TTTAAAATTTTTGGACCCCATA
	pg61-2	pgM9-2	GCTCTCCACTTCACAATTTGG	TGCAGAAGAGTTAGTGCTCCA
	pg61-3	pgM9-3	TGATTTCCAAAGAAATTCCACA	AAGAAGCGGAGAATTTTGTCA
pg63	pg63-1	pgm63-1(7)	GCACATAGCCAGATTACTAACCA	TTCAGATGTGCCTCTAAATGG
pg68	pg68-1	pgM10-1	TTTTGCAATTCACCAGCAAC	GAGGTTCCTACATCATCTGTGG
	pg68-2	pgM10-2	TTTTGCAATTCACCAGCAAC	GGGATAACTGGCCTTGTGAG
	pg68-3	pgM10-3	GCACTTGCAATTGAATCAAAG	AAGACGAGGATAGTCAATCATATCAA
	pg68-4	pgM10-4	TAGGCCACTGCGCTATTGTA	TTCCATGCGAGAAAGACTCA
pg73	pg73-1	pgm73-1(8)	GGAGCAATAGGAGGTAAACCA	CAGCACCCGTGTCAACTAAG
pg75	pg75-1	pgm75-1-26	TGTTGAGCCGGTTTGTATTG	TTCTGCCATAATGACTTGTGAC

4. PCR amplification of MITE containing sequence

PCR amplifications were performed in a total volume of 20 μL containing 20 ng of DNA, 2.5 mM of each dNTP, 10 pmol of each primers, and 0.5 units *Taq* DNA polymerase (Vivagen, Korea) using a DNA Engine Thermal Cycler (Bio-rad, USA). The PCR cycling conditions were as follows: 5 min at 94℃ for denaturation, 35 cycles of 30 sec at 94℃, 30 sec at Tm°C, 1 min at 72℃, and 10 min at 72℃ for final extension. PCR products were separated on 1.5% agarose gels and stained with ethidium bromide (Ameresco, USA).

5. Sequence analyses and comparison

PCR amplicons showed MITE insertion and non-insertion at the expected size were purified by using PCR product purification kit (Solgent, Korea) for DNA sequencing. PCR amplicons were sequenced by ABI3730XL auto-sequencer serviced by NICEM (National Instrumentation for Environment Management, Seoul, Korea).

Genomic sequences were retrieved from *P. ginseng* genome sequence database and then compared by using pipmaker (http://pipmaker.bx.psu.edu/cgi-bin/pipmaker?basic), ClustalW (http://clustalw.ddbj.nig.ac.jp/), and MEGA 5.0 software.

RESULTS

1. AFLP analysis between two *P. ginseng* cultivars

To investigate AFLP present between Chunpoong (CP) and Yunpoong (YP), PCR amplification was conducted using the genomic DNAs of both cultivars. Total 137 primer pairs from 3 x 3 primer combination for *Eco*RI and *Mse*I amplified average 42.4 PCR bands and generated average 1.7 polymorphic bands of 3.9 %. Of 137 primer pairs, 61 could be scored. All of 8 primer pairs from 3 x 4 primer combination could be scored. These 8 primer pairs generated average 29.9 PCR bands with average 1.3 polymorphic bands (4.2%) (Table 3). This result indicates addition of one nucleotide for Mse1 primer increased primer specificity and thus generated less PCR band than that of 3 x 3 primer combination.

Among primer pairs, 10 of 3 x 3 primer combination showing high percentage of polymorphic bands were used for further AFLP analysis of 11 ginseng samples.

Table 3. Result of CP-YP in AFLP amplification

Primer combination (EcoR1 x Mse1)	Total number of primer combination	Scorable primer combination	Average bands per primer combination	Average polymorphic bands per primer combination	%
3 x 3	137	61	42.4	1.7	3.9
3 x 4	8	8	29.9	1.3	4.2

2. AFLP analysis of 11 ginseng samples

The 10 pairs of primer combinations were used for AFLP analysis of 11 ginseng samples and all of them could detect polymorphism.

In addition, 11 ginseng samples showed higher average percentage of polymorphic bands (22.2 %) than that of CP-YP AFLP analysis (3.9%), mainly because there are many polymorphisms between *P. ginseng* cultivars and American ginseng (Table 4). Representative results of AFLP analysis were shown in Figure 1, which showed many polymorphic bands between *P. ginseng* cultivars and American ginseng amplified by No. 2 and No. 3 primer combination.

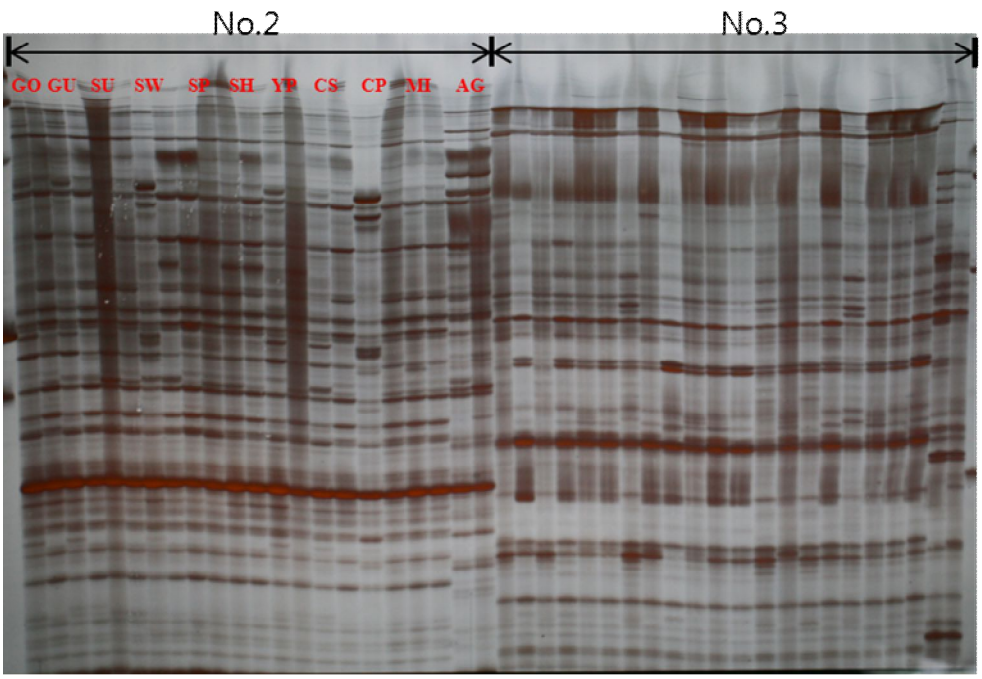

Figure 1. AFLP analysis of 11 ginseng samples by using No. 2 and No. 3 primer combinations. GO : Gopoong, GU : Gumpoong, SU : Sunun, SW : Sunwon, SP : Sunpoong, SH : Sunhyang, YP : Yunpoong, CS : Chungsun, CP : Chunpoong, MI : Mimaki, AG : American ginseng.

Table 4. Summary of AFLP analysis of 11 ginseng samples using the 10 primer pairs

No.	EcoR1	Mse1	CP-YP scored 3 x 3 combination total bands No.	CP-YP scored 3 x 3 combination polymorphic bands No.	Total bands No. in 11 samples	No. of Polymorphic bands	% of polymorphic bands
1	AGG	GGA	not counted	not counted	51	9	17.6
2	AAC	GAT	45	6	76	20	26.3
3	ATG	GAT	39	4	54	14	25.9
4	AAG	GCG	34	5	41	6	14.6
5	AAG	GCC	61	7	56	8	14.3
6	ATT	GCT	46	6	59	12	20.3
7	ATG	GCT	50	3	47	10	21.3
8	ATC	GCT	43	2	51	12	23.5
9	ATC	GGC	39	3	51	12	23.5
10	ATG	GGT	49	2	63	19	30.2
Average			**45.1**	**4.2**	**54.9**	**12.2**	**22.2**

Five specific bands could be distinguished among 11 ginseng samples (Table 5), indicating that they can be used as the molecular markers for identification of *P. ginseng* cultivars. For example, in the No. 9 primer combination, 6th band at 2,100 bp can be a marker for GU, SW and CP, and 35th band at 280 bp for Mi.

Table 5. Summary of specific bands to be used for molecular marker

NO.	EcoR1	Mse1	specific bands among 11 samples
1	AGG	GGA	17th band for SP
2	AAC	GAT	30th band for GU and SW
6	ATT	GCT	49th band for SU, SW, YP and AG
9	ATC	GGC	6th band for GU, SW and CP 35th band for Mi

Taken together, total 549 bands were amplified by AFLP analysis of 11 ginseng samples with 10 primer combinations. Among them, 117 were polymorphic between *P. ginseng* and *P. quinquefolium*, and 5 were polymorphic among *P. ginseng* cultivars (Table 6).

Table 6. Summary of AFLP-based polymorphism identified in this study

Contents	No. of bands	PPB (%)
Total in 11 ginseng samples	549	-
Polymorphic between species	117	21.3
Polymorphic among *P. ginseng* cv	5	0.9

PPB : Percentage of polymorphic bands

3. Identification of MITE consensus sequences

Work flow to identify MITE consensus sequences is shown in Figure 2. Using *P. ginseng* whole genome draft sequence generated by NGS technique, 133 MITE consensus sequences in *P. ginseng* scaffold were identified by the program MITE-Hunter. After determining intact TIRs by Einverted software, 78 of 133 MITE consensus sequences were chosen. Finally 49 MITE families were selected by 1st and 2nd selection. Afterward, proper scaffold sequences with MITE were selected by Blast search using MITE consensus sequences as queries and then used to design PCR primers to amplify region with MITE. As a result, 73 scaffolds of 25 families were selected and analyzed further (Table 7). Additionally, hypothetical hairpin loop structures of the 25 MITE consensus sequences were identified by MFOLD software (Figure 3).

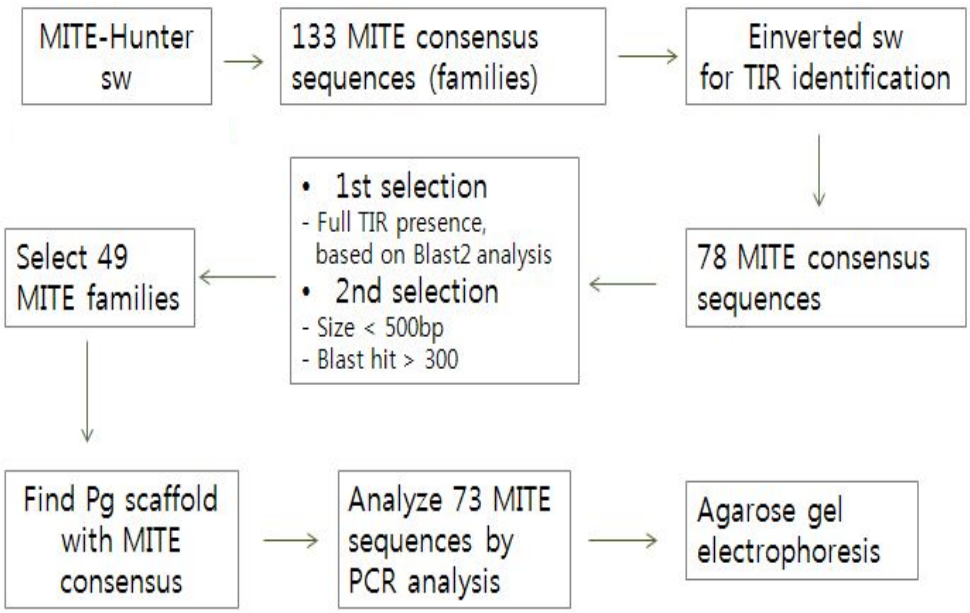

Figure 2. Work flow for development MITE-based markers

Table 7. Summary of 78 MITE families found in *P. ginseng* scaffold database

Family	SIZE (bp)	Blast Hits	1st, 2nd selection	Numbers of scaffolds used for MIP
PG1	183	1168	o	
PG2	283	712	o	6
PG3	108	641	o	1
PG4	267	652	o	3
PG5	116	80		
PG6	403	341	o	1
PG7	322	452	o	4
PG8	350	464	o	
PG9	299	1173	o	
PG10	248	268		
PG11	147	1256	o	
PG12	208	629	o	2
PG13	166	1347	o	
PG14	375	559	o	8
PG15	1240	139		
PG16	301	126	o	4
PG17	261	585	o	7
PG18	861	45		
PG19	120	605	o	
PG20	216	765	o	1
PG21	154	634	o	
PG22	799	469		
PG23	347	250		
PG24	255	525	o	4
PG25	198	148		

Table 7. (Continued)

Family	SIZE (bp)	Blast Hits	1st, 2nd selection	Numbers of scaffolds used for MIP
PG26	260	156	o	4
PG27	529	899		
PG28	264	1150	o	
PG29	699	236		
PG30	96	28		
PG31	183	242		
PG32	1327	2359		
PG33	692	585		
PG34	345	1075	o	2
PG35	636	202		
PG36	137	1393	o	1
PG37	339	724	o	1
PG38	345	31		
PG39	295	633	o	
PG40	878	1091		
PG41	335	903	o	
PG42	108	560	o	2
PG43	615	181		
PG44	297	585	o	
PG45	319	1546	o	
PG46	277	189	o	4
PG47	385	68	o	5
PG48	238	907	o	1
PG49	231	60		
PG50	201	327	o	
PG51	244	584	o	2
PG52	263	492	o	

Table 7. (Continued)

Family	SIZE (bp)	Blast Hits	1st, 2nd selection	Numbers of scaffolds used for MIP
PG53	228	304	o	
PG54	287	1837	o	
PG55	366	74		
PG56	167	69		
PG57	112	77		
PG58	222	234		
PG59	378	396	o	
PG60	414	1425	o	
PG61	181	100	o	3
PG62	1134	460		
PG63	382	421	o	1
PG64	132	60		
PG65	388	248		
PG66	409	1015	o	
PG67	192	321	o	
PG68	290	30	o	4
PG69	922	45		
PG70	273	593	o	
PG71	292	697	o	
PG72	328	570	o	
PG73	189	918	o	1
PG74	1254	316		
PG75	380	452	o	1
PG76	259	158		
PG77	194	765	o	
PG78	573	765		
78			**49 families**	**73 (in 25 families)**

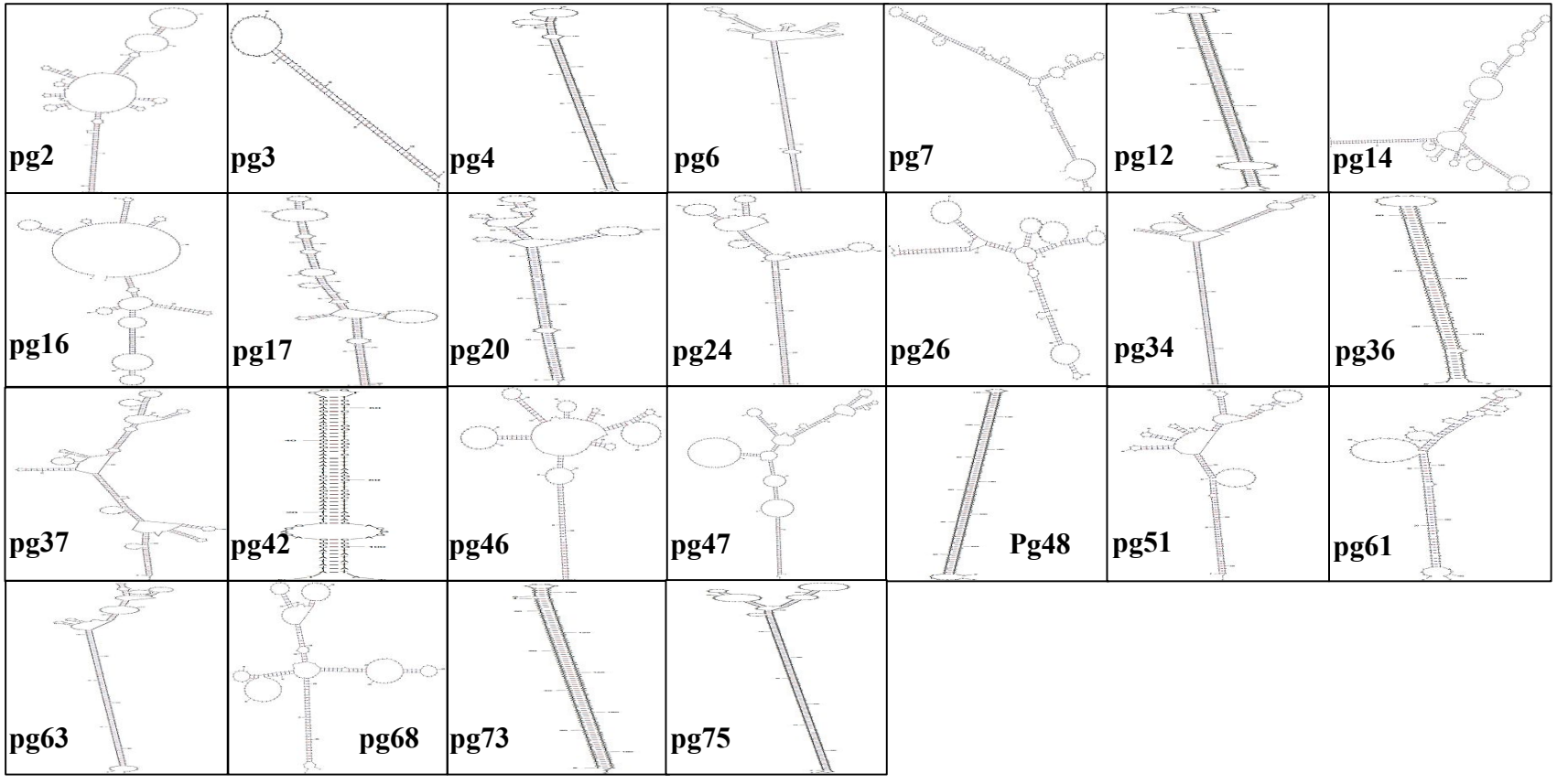

Figure 3. Hypothetical secondary structure of 25 MITE consensus sequences predicted by MFOLD software

4. Analysis of 73 scaffolds with MITE

MITE structure present in the selected 73 scaffolds was analysed, as shown in Table 8. Of MITE sequences in 73 scaffolds, 42 were *stowaway*-like (TSD: TA) MITE and the rest were other MITE type. Especially pg3-1 was *Gaijin*-like MITE which has 3-bp TSDs (5′-TGA-3′) and 41bp TIRs, pg24-2 was *Mutator*-like MITE which has 9-bp TSDs (5′-TACTATTTA-3′) and 74 bp TIRs, and 26 members were *Explorer*-like MITE which has no TSDs [32, 33].

Pg37-1 had shortest TIRs (17 bp) and pg75-1 had the longest TIRs (157 bp) and almost all of the MITE sequence was TIRs, pg3-1was smallest MITE (107 bp) and pg47-1 was the longest MITE (389 bp).

Table 8. Identification of MITE elements in *P. ginseng* 73 scaffold sequences

MITE family	Members	Scaffold	Scaffold length (bp)	Primer	TSD	TIR (consensus sequence)	MITE size (bp)	Predicted amplicon size (bp)
pg2	pg2-1	scaffold65243.1	7357	pgM1-1	TA	28 bp CTCCCTCCATTCCTAAA TAAGTGA…	291	628
	pg2-2	scaffold207055.1	1015	pgM1-2	TA		288	601
	pg2-3	scaffold3641.1	74160	pgM1-3	TA		282	723
	pg2-4	scaffold93892.1	2667	pgM1-4	TA		288	672
	pg2-5	scaffold75499.1	5480	pgm02-1(1)	TA		288	601
	pg2-6	scaffold33536.1	20806	pgm02-2(2)	TA		288	606
pg3	pg3-1	scaffold1534.1	96886	pgm3-1-1	TGA	41 bp CTGTTGGGATTTCACAT GAATCCCAA…	107	406
pg4	pg4-1	scaffold86544.1	3884	pgm4-1-2	TTAAAAT	104 bp TTAAACCCCGAAGTATA CACTCTTT...	266	500
	pg4-2	scaffold2797.1	79445	pgm4-2-3	TA		266	613
	pg4-3	scaffold28.1	213813	pgm4-3-4	TA		266	619
pg6	pg6-1	scaffold1984.1	88605	pgm6-1-5	TA	127 bp ACTGCAAAAGTGTCCCT CAAAGTT…	400	728
pg7	pg7-1	scaffold76132.1	5389	pgM2-1	-	60 bp TCCAATTTCATCCAATA CACCATA…	323	816
	pg7-2	scaffold49040.1	12626	pgM2-2	-		322	888
	pg7-3	scaffold12857.1	42172	pgM2-3	-		257	561
	pg7-4	scaffold1147.1	103130	pgM2-4	-		255	615

Table 8. (Continued)

MITE family	Members	Scaffold	Scaffold length (bp)	Primer	TSD	TIR (consensus sequence)	MITE size (bp)	Predicted amplicon size (bp)
pg12	pg12-1	scaffold1917.1	89478	pgm12-1-6	TA	91 bp CGGAGAGGCTCTAAAG CGAACCTTAA…	207	510
	pg12-2	scaffold6580.1	57389	pgm12-2-7	TATA		207	531
pg14	pg14-1	scaffold525.1	125089	pgM3-1	TA	32 bp CTCCCTCCGTCCAAGTT TACTTGTCCT…	389	744
	pg14-2	scaffold9712.1	49207	pgM3-2	TA		384	729
	pg14-3	scaffold17121.1	36631	pgM3-3	-		384	779
	pg14-4	scaffold10639.1	46841	pgM3-4	-		384	776
	pg14-5	scaffold13378.1	40661	pgm14-1-8	TA		375	700
	pg14-6	scaffold4946.1	65619	pgm14-2-9	TA		374	701
	pg14-7	scaffold6440.1	57980	pgm14-3-10	TA		375	709
	pg14-8	scaffold1793.1	94378	pgm14-4-11	TA		365	642
pg16	pg16-1	scaffold49418.1	12669	pgM4-1	TA	33 bp TTCCCTCCTTTTCATTAT GTAGGTCG…	298	601
	pg16-2	scaffold22944.1	28976	pgM4-2	TA		296	632
	pg16-3	scaffold26388.1	25710	pgM4-3	TA		282	609
	pg16-4	scaffold37742.1	18142	pgM4-4	TA		281	600

Table 8. (Continued)

MITE family	Members	Scaffold	Scaffold length (bp)	Primer	TSD	TIR (consensus sequence)	MITE size (bp)	Predicted amplicon size (bp)
pg17	pg17-1	scaffold5819.1	63945	pgM5-2	-	39 bp CTCCCTCCGTCCTCAAAAGG…	266	629
	pg17-2	scaffold16281.1	37598	pgM5-3	TA		267	616
	pg17-3	scaffold22691.1	29438	pgM5-4	-		266	658
	pg17-4	scaffold6732.1	60064	pgm17-1-12	TA		265	600
	pg17-5	scaffold26344.1	26495	pgm17-2-13	TA		261	600
	pg17-6	scaffold126895.1	1608	pgm17-3-14	TA		257	510
	pg17-7	scaffold100818.1	2208	pgm17-4-15	TA		260	601
pg20	pg20-1	scaffold12957.1	43619	pgm20-1-16	TA	64 bp TTAGAGATACATTTTAATTTT…	208	529
pg24	pg24-1	scaffold22132.1	30168	pgm24-1-17	TA	74 bp CTCCCTCCGTCCAGCTTTAT…	252	600
	pg24-2	scaffold24462.1	28349	pgm24-2-18	TACTATTTA		254	600
	pg24-3	scaffold27388.1	25824	pgm24-3-19	TA		247	500
	pg24-4	scaffold30383.1	23000	pgm24-4-20	TA		253	501
pg26	pg26-1	scaffold36110.1	19090	pgM6-1	-	20 bp CCTCCATCCCAAATTATCTG	266	602
	pg26-2	scaffold30066.1	23956	pgM6-2	-		259	600
	pg26-3	scaffold82192.1	4470	pgM6-3	-		257	525

Table 8. (Continued)

MITE family	Members	Scaffold	Scaffold length (bp)	Primer	TSD	TIR (consensus sequence)	MITE size (bp)	Predicted amplicon size (bp)
pg26	pg26-4	scaffold30246.1	22858	pgM6-4	-	20 bp CCTCCATCCCAAATTATCTG	249	639
pg34	pg34-1	scaffold27553.1	25466	pgm34-1(3)	-	106 bp GGCTATGTTTGTTTGCAAAAAATATG…	344	700
	pg34-2	scaffold689.1	116830	pgm34-2(4)	TA		342	606
pg36	pg36-1	scaffold16975.1	35757	pgm36-1(5)	TA	53 bp TGGGGTTCCATCCACAGCAAC…	124	412
pg37	pg37-1	scaffold15838.1	37957	pgm37-1-21	-	17 bp AGACTATATTTAGTCTA	337	602
pg42	pg42-1	scaffold4817.1	66269	pgm42-1-22	-	44 bp CGGAGAGGCTCTAAAGCGAAC…	107	352
	pg42-2	scaffold937.1	108402	pgm42-2-23	TA		107	353
pg46	pg46-1	scaffold3387.1	75881	pgM7-1	TA	34 bp CTCCCTCCGTCCCAAAAAAGATGA…	281	637
	pg46-2	scaffold4892.1	66866	pgM7-2	TA		278	634
	pg46-3	scaffold5219.1	64217	pgM7-3	-		276	664
	pg46-4	scaffold12055.1	44794	pgM7-4	-		272	603
pg47	pg47-1	scaffold15492.1	37926	pgM8-1	TA	33 bp CTCCCTCCGTCCCATATTAAAC…	389	772
	pg47-2	scaffold36525.1	18938	pgM8-2	TA		384	747
	pg47-3	scaffold12604.1	43397	pgM8-3	TA		381	749
	pg47-4	scaffold2752.1	82761	pgM8-4	-		384	704

Table 8. (Continued)

MITE family	Members	Scaffold	Scaffold length (bp)	Primer	TSD	TIR (consensus sequence)	MITE size (bp)	Predicted amplicon size (bp)
pg47	pg47-5	scaffold8339.1	55035	pgM8-5	-	33 bp ACTCCCTCCGTCCCATATTAAA…	377	712
pg48	pg48-1	scaffold1732.1	91548	pgm48-1(6)	-	65 bp GTGATTAGGCTGTTGCATGATA…	211	603
pg51	pg51-1	scaffold2256.1	86009	pgm51-1-24	TA	41 bp CTCCCTCTGTTCCACAAAAAGA…	243	508
	pg51-2	scaffold2999.1	79363	pgm51-2-25	TA		243	520
pg61	pg61-1	scaffold7357.1	55594	pgM9-1	TA	45 bp GAATAATTATATGTGCACAAATTA…	179	500
	pg61-2	scaffold15082.1	38789	pgM9-2	-		173	584
	pg61-3	scaffold11451.1	45460	pgM9-3	-		172	505
pg63	pg63-1	scaffold104886.1	2019	pgm63-1(7)	TATA	120 bp GGGGGCGGATCAATGACACTA…	370	728
pg68	pg68-1	scaffold12239.1	44409	pgM10-1	-	30 bp CTCCCTCCATTTCATTTTAATAGA…	293	701
	pg68-2	scaffold4578.1	67597	pgM10-2	-		293	639
	pg68-3	scaffold40055.1	16669	pgM10-3	-		207	563
	pg68-4	scaffold951.1	109093	pgM10-4	-		205	504
pg73	pg73-1	scaffold15132.1	38367	pgm73-1(8)	TA	86 bp TGTATGGGGATGGTGATCCATC…	198	501
pg75	pg75-1	scaffold9158.1	50974	pgm75-1-26	TA	157 bp AGATTGACCTAAGATAAGTTAT…	371	710

5. PCR analysis of MITE-based insertional polymorphism

With primers designed based on 73 scaffold sequences with MITE, MITE-based insertional polymorphism (MIP) was analysed and summarized, as shown in Table 9 and 10. Total 169 bands were amplified with an average 2.3 bands per primers. Among 73 PCR primer markers, 16 (21.9%) generated polymorphic bands between *P. ginseng* (*Pg*) and *P. quinquefolium* (*Pq*). Of them, 13 did not show MIP between *Pg* and *Pq*, whereas 3 showed MIP between both species. Figure 4 and 5 show analysis of MITE derived from scaffold76132 without MIP and scaffold86544 with MIP, respectively.

Among 73 primer pairs for PCR analysis of MITE, 10 generated MIP bands, of which 4 amplified 2 bands and rest 6 more than 3 bands (Table 11). The MITE flanking sequences with the 10 primer site were further analysed at nucleotide level.

Table 9. Summary of MITE-PCR analysis (* indicates result of MIP bands)

MITE family	Members	Scaffold	Primer	Polymorphic Between	No. of amplicon
pg2	pg2-1*	scaffold65243.1	pgM1-1	Pg vs. Pq	≥ 3
	pg2-2	scaffold207055.1	pgM1-2	no	1
	pg2-3	scaffold3641.1	pgM1-3	no	1
	pg2-4	scaffold93892.1	pgM1-4	Pg vs. Pq	2
	pg2-5	scaffold75499.1	pgm02-1(1)	Pg vs. Pq	2
	pg2-6	scaffold33536.1	pgm02-2(2)	no	1
pg3	pg3-1	scaffold1534.1	pgm3-1-1	Pg vs. Pq	≥ 3
pg4	pg4-1*	scaffold86544.1	pgm4-1-2	Pg vs. Pq	2
	pg4-2	scaffold2797.1	pgm4-2-3	no	3
	pg4-3*	scaffold28.1	pgm4-3-4	no	≥ 3
pg6	pg6-1	scaffold1984.1	pgm6-1-5	no	≥ 3
pg7	pg7-1	scaffold76132.1	pgM2-1	Pg vs. Pq	≥ 3
	pg7-2	scaffold49040.1	pgM2-2	Pg vs. Pq	2
	pg7-3	scaffold12857.1	pgM2-3	no	≥ 3
	pg7-4	scaffold1147.1	pgM2-4	no	1
pg12	pg12-1	scaffold1917.1	pgm12-1-6	no	≥ 3
	pg12-2	scaffold6580.1	pgm12-2-7	no	≥ 3
pg14	pg14-1*	scaffold525.1	pgM3-1	no	2
	pg14-2	scaffold9712.1	pgM3-2	no	1

Table 9. (Continued)

MITE family	Members	Scaffold	Primer	Polymorphic Between	No. of amplicon
pg14	pg14-3	scaffold17121.1	pgM3-3	no	≥ 3
	pg14-4*	scaffold10639.1	pgM3-4	no	2
	pg14-5	scaffold13378.1	pgm14-1-8	no	1
	pg14-6	scaffold4946.1	pgm14-2-9	no	2
	pg14-7	scaffold6440.1	pgm14-3-10	no	1
	pg14-8	scaffold1793.1	pgm14-4-11	no	2
pg16	pg16-1	scaffold49418.1	pgM4-1	no	1
	pg16-2	scaffold22944.1	pgM4-2	no	2
	pg16-3*	scaffold26388.1	pgM4-3	no	2
	pg16-4	scaffold37742.1	pgM4-4	no	1
pg17	pg17-1	scaffold5819.1	pgM5-2	no	2
	pg17-2	scaffold16281.1	pgM5-3	no	1
	pg17-3	scaffold22691.1	pgM5-4	no	1
	pg17-4	scaffold6732.1	pgm17-1-12	no	≥ 3
	pg17-5	scaffold26344.1	pgm17-2-13	no	2
	pg17-6	scaffold126895.1	pgm17-3-14	no	2
	pg17-7*	scaffold100818.1	pgm17-4-15	no	≥ 3
pg20	pg20-1*	scaffold12957.1	pgm20-1-16	no	≥ 3

Table 9. (Continued)

MITE family	Members	Scaffold	Primer	Polymorphic Between	No. of amplicon
pg24	pg24-1	scaffold22132.1	pgm24-1-17	no	1
	pg24-2*	scaffold24462.1	pgm24-2-18	no	≥ 3
	pg24-3	scaffold27388.1	pgm24-3-19	no	1
	pg24-4	scaffold30383.1	pgm24-4-20	Pg vs. Pq	2
pg26	pg26-1	scaffold36110.1	pgM6-1	no	2
	pg26-2	scaffold30066.1	pgM6-2	no	≥ 3
	pg26-3	scaffold82192.1	pgM6-3	no	2
	pg26-4	scaffold30246.1	pgM6-4	no	2
pg34	pg34-1	scaffold27553.1	pgm34-1(3)	no	2
	pg34-2	scaffold689.1	pgm34-2(4)	no	≥ 3
pg36	pg36-1	scaffold16975.1	pgm36-1(5)	no	≥ 3
pg37	pg37-1	scaffold15838.1	pgm37-1-21	no	1
pg42	pg42-1	scaffold4817.1	pgm42-1-22	Pg vs. Pq	2
	pg42-2	scaffold937.1	pgm42-2-23	Pg vs. Pq	≥ 3
pg46	pg46-1	scaffold3387.1	pgM7-1	no	1
	pg46-2	scaffold4892.1	pgM7-2	no	1
	pg46-3	scaffold5219.1	pgM7-3	no	≥ 3
	pg46-4	scaffold12055.1	pgM7-4	Pg vs. Pq	2

Table 9. (Continued)

MITE family	Members	Scaffold	Primer	Polymorphic Between	No. of amplicon
pg47	pg47-1*	scaffold15492.1	pgM8-1	Pg vs. Pq	≥ 3
	pg47-2	scaffold36525.1	pgM8-2	no	≥ 3
	pg47-3	scaffold12604.1	pgM8-3	no	≥ 3
	pg47-4	scaffold2752.1	pgM8-4	no	≥ 3
	pg47-5	scaffold8339.1	pgM8-5	no	1
pg48	pg48-1	scaffold1732.1	pgm48-1(6)	no	1
pg51	pg51-1	scaffold2256.1	pgm51-1-24	Pg vs. Pq	2
	pg51-2	scaffold2999.1	pgm51-2-25	no	2
pg61	pg61-1	scaffold7357.1	pgM9-1	no	1
	pg61-2	scaffold15082.1	pgM9-2	no	≥ 3
	pg61-3	scaffold11451.1	pgM9-3	no	1
pg63	pg63-1	scaffold104886.1	pgm63-1(7)	no	1
pg68	pg68-1	scaffold12239.1	pgM10-1	no	2
	pg68-2	scaffold4578.1	pgM10-2	no	≥ 3
	pg68-3	scaffold40055.1	pgM10-3	Pg vs. Pq	1
	pg68-4	scaffold951.1	pgM10-4	Pg vs. Pq	1
pg73	pg73-1	scaffold15132.1	pgm73-1(8)	no	≥ 3
pg75	pg75-1	scaffold9158.1	pgm75-1-26	Pg vs. Pq	≥ 3
25 families		**73 scaffolds**		**16 (Pg vs. Pq)**	**169 bands (average 2.3 bands)**

Table 10. Classification of bands based on polymorphism

No. of amplicon	no polymorphism	polymorphic between Pg vs. Pq	MIP between Pg vs. Pq	Total
1	21	2	-	23
2	15	7	1	23
≥ 3	21	4	2	27
total	**57**	**13**	**3**	**73**

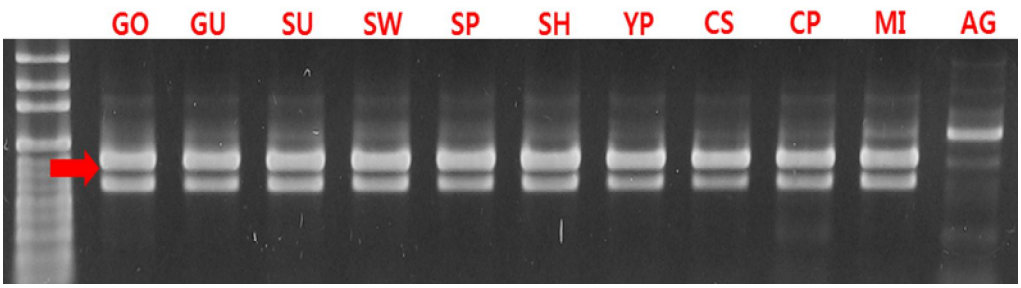

Figure 4. Polymorphic DNA band between *P. ginseng* and *P. quinquefolium,* when pg7-type MITE sequence present in scaffold76132 was amplified. Arrow indicates predicted 816 bp of amplicon with MITE.

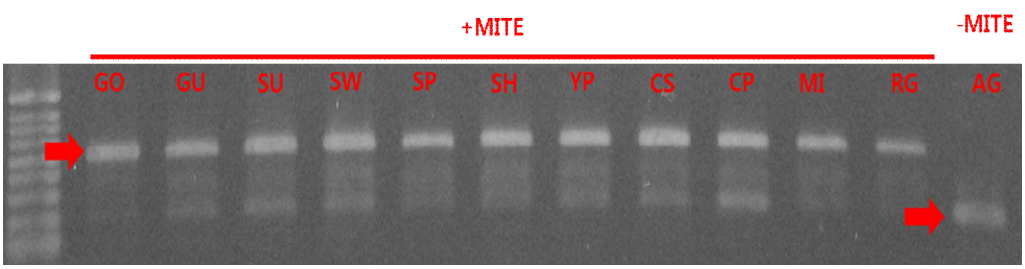

Figure 5. MIP between *P. ginseng* and *P. quinquefolium*, when pg4-type MITE sequence present in scaffold86544 was amplified. Arrows indicate predicted 500 bp and 234 bp of amplicons with or without MITE, respectively.

Table 11. Classification of MIP bands based on amplicon numbers

No. of amplicon	MIP between bands	No MIP	Total
1	-	23	23
2	4	19	23
≥ 3	6	21	27
total	**10**	**63**	**73**

6. Sequence comparison of MITE flanking sequences showing MIP

Of 10 MITE sequences showing MIP, pg24-type MITE sequence present in scaffold24462 generated two amplicons, when performed PCR analysis using primers that could amplify regions with the MITE (Figure 6A). PCR band no. 11-1 and no. 12-1 were amplified from templates of *P. ginseng* cultivars and *P. quinquefolium* respectively, and also both bands were considered as DNA fragments with MITE insertion. On the other hand, PCR band no.12-2 from template of *P. quinquefolium* was considered to have no MITE insert.

To identify and compare MITE-flanking sequences, the three PCR bands were purified and sequenced. Through analysis of nucleotide sequences, no.11-1 and no. 12-1 DNA fragments were identified to have MITE sequence, whereas no. 12-2 was not, as expected based on band size. Nucleotide sequence with MITE of no. 11-1 is shown in Figure 6B. The other two PCR bands were also the same as no. 11-1 in nucleotide sequence except MITE insertion region. When searched *P. ginseng* scaffolds by

BlastN using nucleotide sequences of the three PCR bands, both no. 11-1 and no. 12-1 were mapped to scaffold24462, while no. 12-2 was mapped to both scaffold28852 and scaffold56235 (Figure 6C). In addition, other regions far from MITE insertion in scaffold24462 were also compared with those in scaffold28852 and scaffold56235 (Figure 7). Flanking sequences of MITE were well conserved among three scaffold sequences, although MITE was present in one scaffold or absent in the other two scaffolds. This indicates that MIP was resulted from the conserved region among these three scaffold sequences.

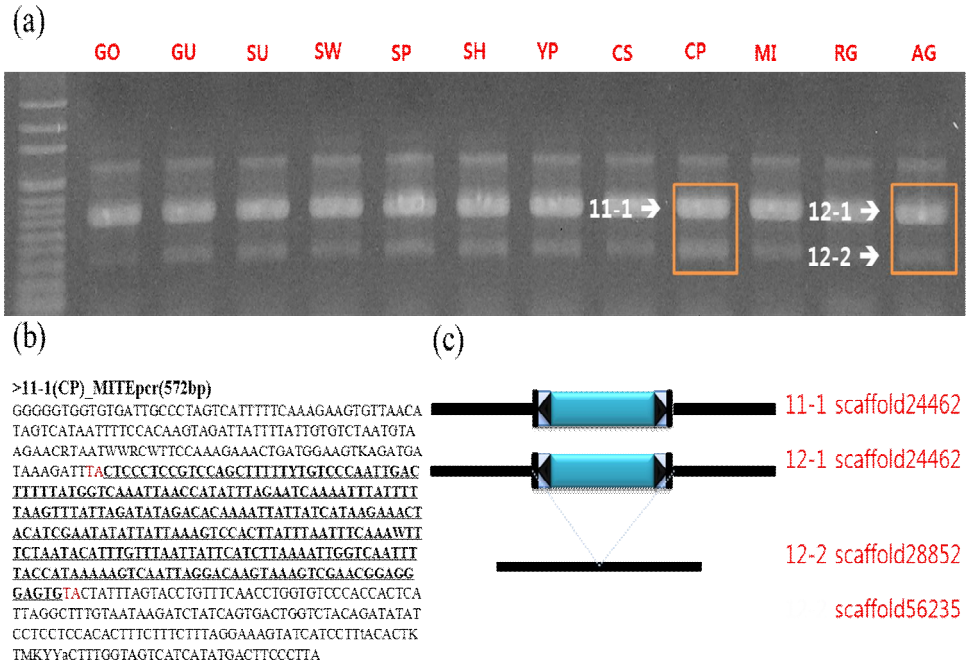

Figure 6. Sequence analyses of pg24-type MITE sequence present in scaffold24462. (a) Three MIP bands, no.11-1, no.12-1, and no.12-2, were purified and sequenced. (b) Nucleotide sequence of no. 11-1 PCR band. MITE sequence is shown as bold underlined letters with flanking 2 bp TSDs on either sides. (c) A hypothetical model for MIP. Both No.11-1 and no. 12-1 sequences with MITE were mapped to scaffold24462, while no.12-2 without MITE was mapped to both scaffold28852 and scaffold56235.

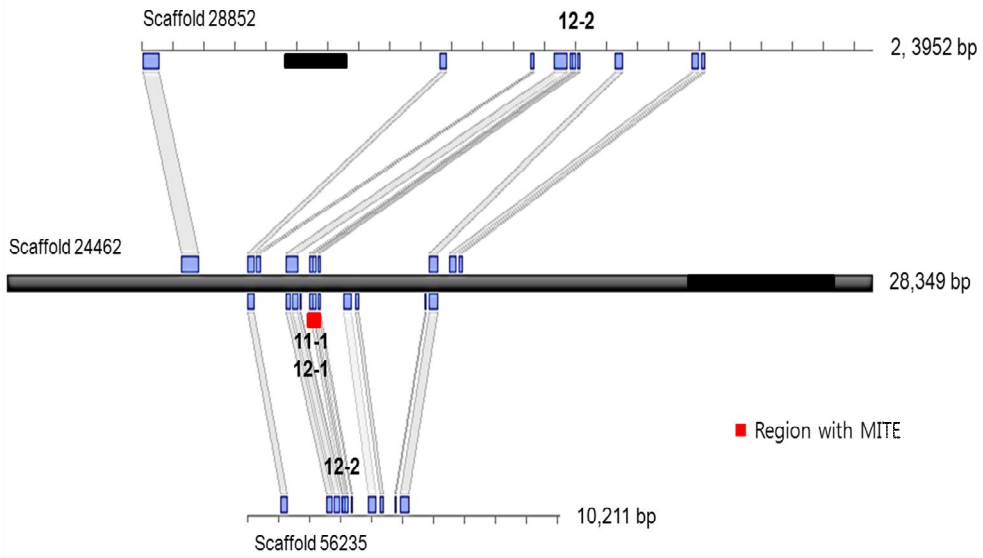

Figure 7. Sequence comparison of three scaffolds, scaffold24462, scaffold28852 and scaffold56235. Boxes linked together indicate conserved sequence between two scaffolds.

To investigate MIP in detail, sequence comparison was conducted by multiple alignments of nucleotide sequences using Pipmaker and ClustalW.

MITE sequence in scaffold24462 showed typical structure of MITE-inserted region, whereas any MITE-related sequence was not found in both scaffold28852 and scaffold56235 (Figure 8). Flanking sequences of MITE were well conserved among the three scaffolds, although there were several InDel and nucleotide substitution (Figure 9a). Meanwhile, these differences in nucleotide sequences were thought to make flanking sequence of MITE in scaffold24462 phylogenetically separated far from those of the other two scaffolds (Figure 9b). In addition, deletion in position of reverse primer was

found in flanking region without MITE, expecting that mutation in primer site might affect amplification efficiency, as shown in Figure 6a.

MITE sequence in scaffold65243 did not show typical structure of MITE-inserted region, when analyzed by Pipmaker (Figure 10). However, typical features of MITE such as TIR and TSD could be found by multiple sequence alignment (Figure 11). As shown in scaffold24462, flanking sequences except MITE of scaffold65243 were well conserved in the other two scaffolds, scaffold932 and scaffold4192 but phylogenetically far from those of the two scaffolds (Figure 11).

MITE analysis of scaffold28 showed MIP band pattern. However, only one scaffold21308 with MITE could be found as a scaffold showing similarity to flanking sequence of MITE in scaffold28, by searching on *P. ginseng* genome sequence database. It may be due to low whole genome coverage of the database. These two scaffolds showed high similarity in MITE sequence as well as in flanking sequences.

In summary, 4 and 5 of 10 MIP sequences could be mapped to 3 and 2 scaffolds, respectively, and the remaining one could be mapped to single scaffold (Table 12).

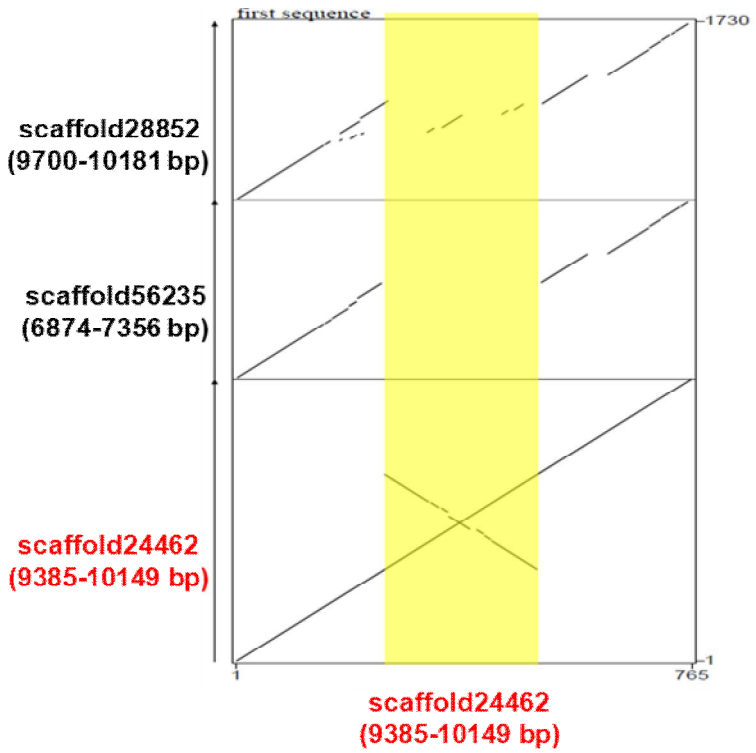

Figure 8. Sequence comparison of flanking regions of MITE in three scaffolds, scaffold24462, scaffold28852 and scaffold56235, using Pipmaker. Middle of the box indicates MITE insertion region.

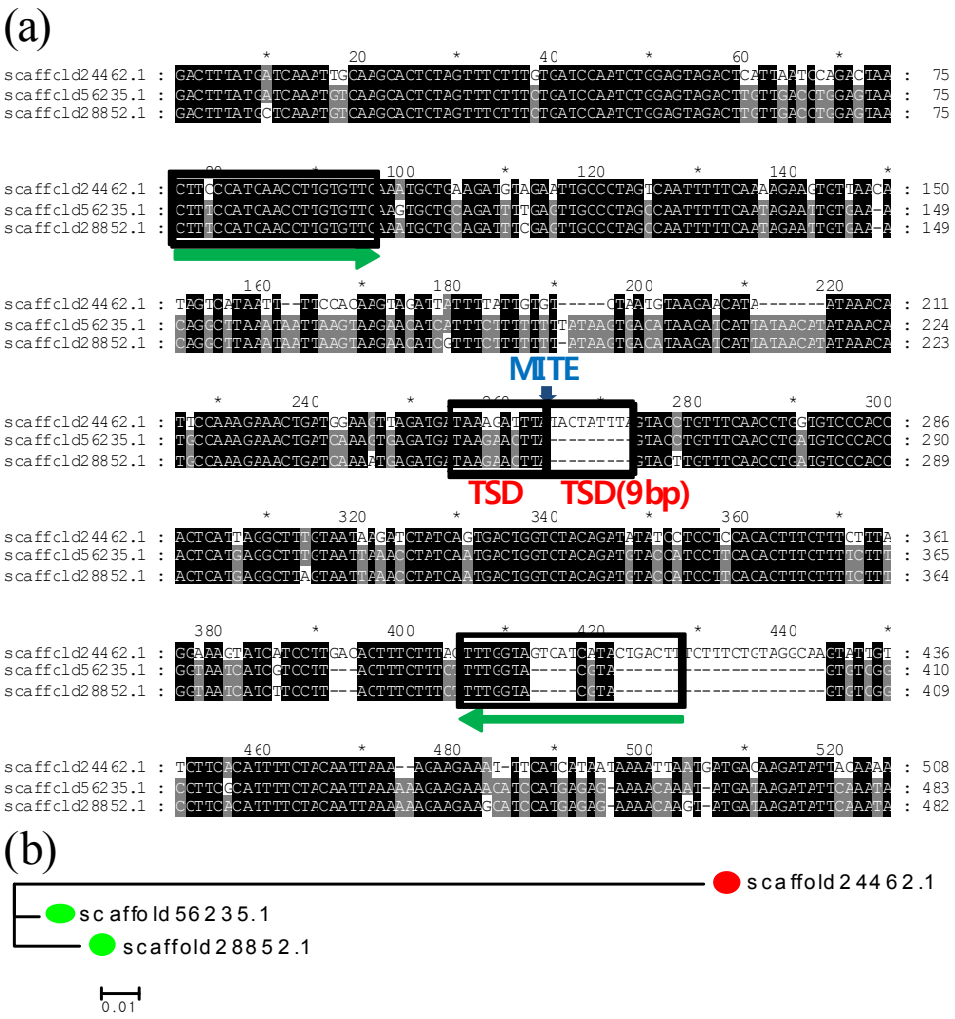

Figure 9. Multiple sequence alignment (a) and phylogenetic tree (b) of flanking region of MITE in three scaffolds. The primer sequences are indicated by boxes and arrows; TSDs by boxes with letter.

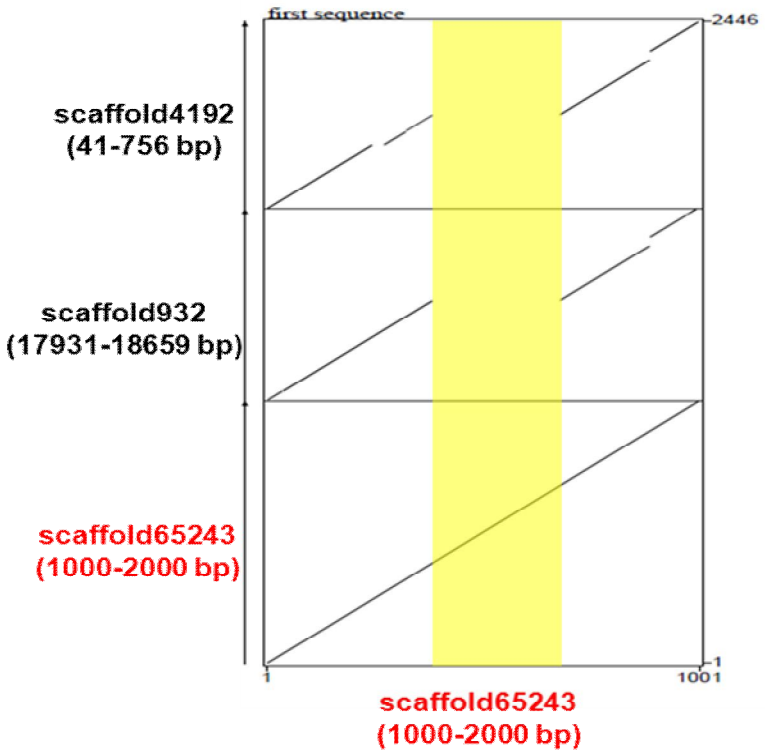

Figure 10. Sequence comparison of flanking regions of MITE in three scaffolds, scaffold4192, scaffold932 and scaffold65243, using Pipmaker. Middle of the box indicates MITE insertion region.

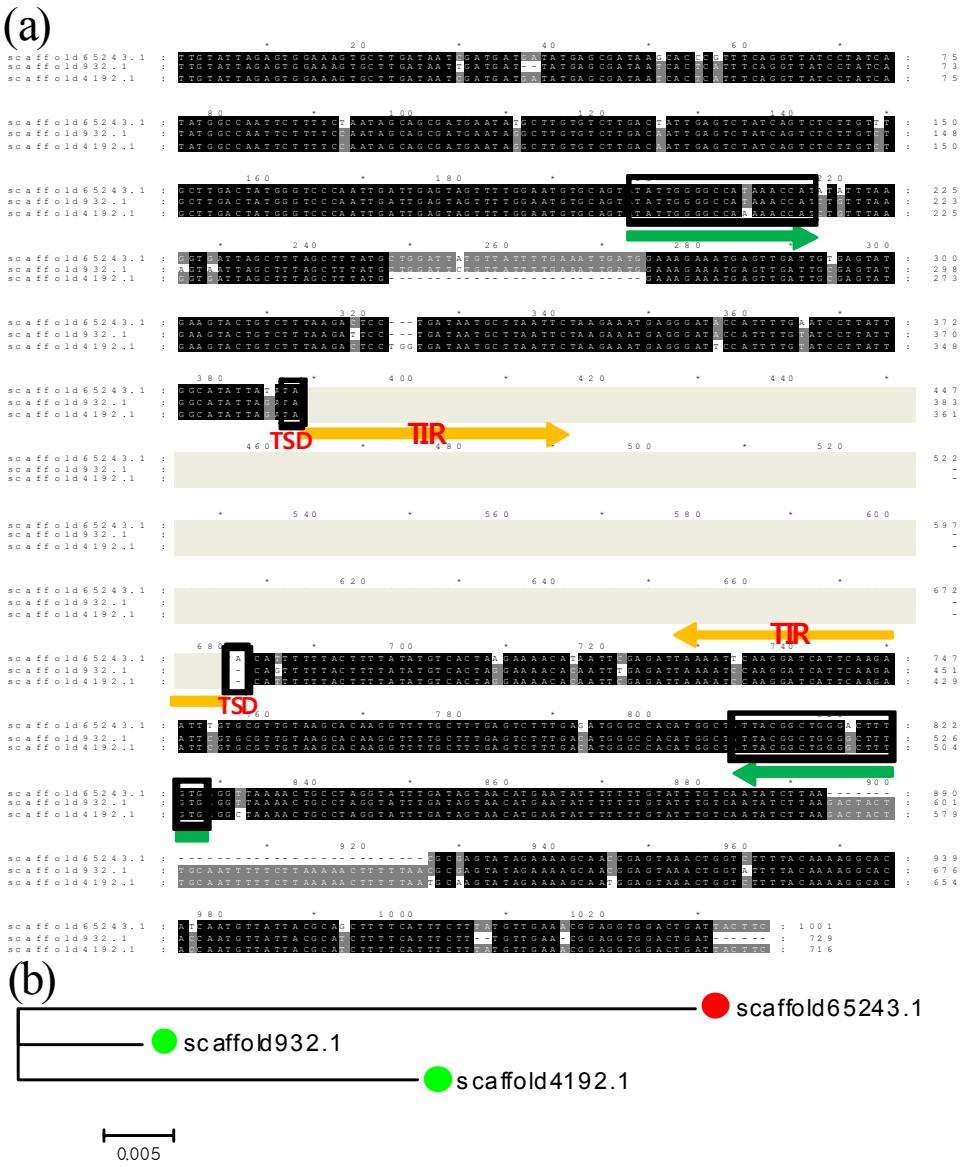

Figure 11. Multiple sequence alignment (a) and phylogenetic tree (b) of flanking region of MITE in three scaffolds. The primer sequences are indicated by boxes and arrows; TSDs and TIRs by boxes with letters; MITE by a light shaded box.

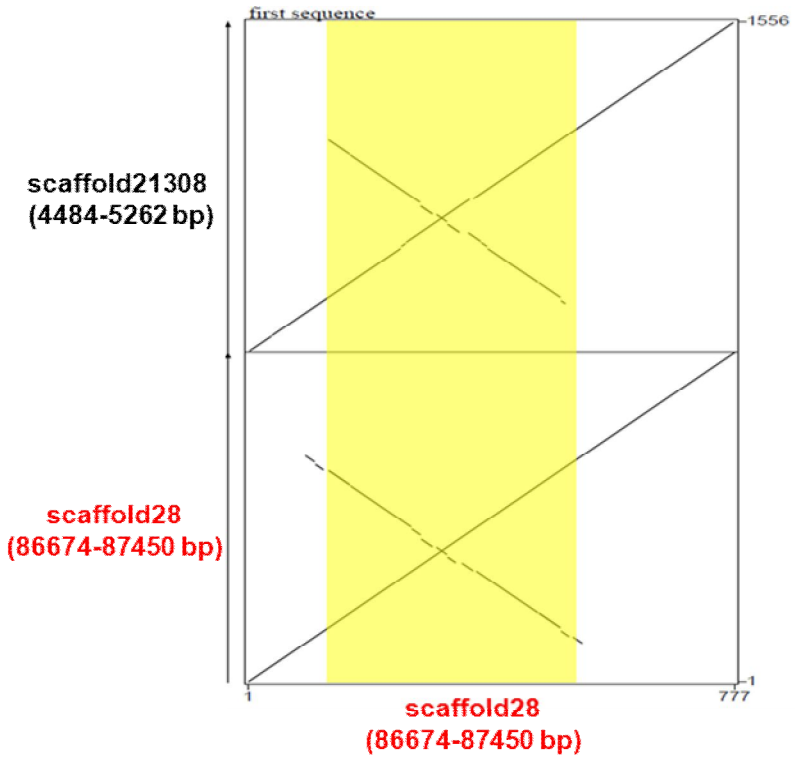

Figure 12. Sequence comparison of flanking regions of MITE in two scaffolds, scaffold28 and scaffold21308, using Pipmaker. Middle of the box indicates MITE insertion region.

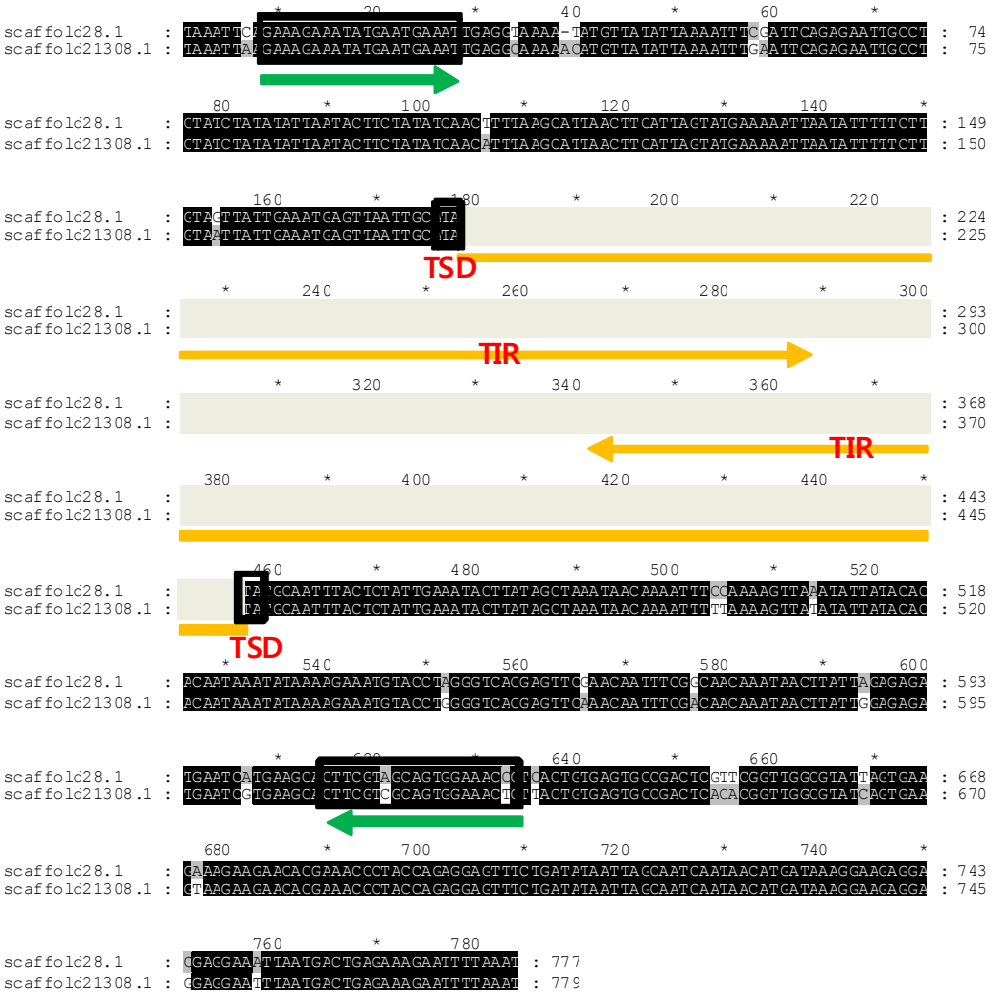

Figure 13. Multiple sequence alignment with flanking region of MITE in two scaffolds. The primer sequences are indicated by boxes and arrows; TSDs by boxes with letters and TIRs by arrows with letters; MITE by a shaded box.

Table 12. List of scaffolds showing similarity in flanking region of MITE

MITE family	Members	Scaffold	Homologous scaffold
pg2	pg2-1	scaffold65243.1(+)	scaffold932.1(-), scaffold4192.1(-)
pg4	pg4-1	scaffold86544.1(+)	scaffold5180.1(-), scaffold23616.1(-)
	pg4-3	scaffold28.1(+)	scaffold21308.1(+)
pg14	pg14-1	scaffold525.1(+)	scaffold1787.1(+), scaffold40987.1(+)
	pg14-4	scaffold10639.1(+)	scaffold17121.1(+)
pg16	pg16-3	scaffold26388.1(+)	-
pg17	pg17-7	scaffold100818.1(+)	scaffold4921.1(+)
pg20	pg20-1	scaffold12957.1(+)	scaffold43642.1(+)
pg24	pg24-2	scaffold24462.1(+)	scaffold28852.1(-), scaffold56235.1(-)
pg47	pg47-1	scaffold15492.1(+)	scaffold2523.1(+)

(+) : with MITE, (-) : without MITE

DISCUSSION

Korean ginseng (*P. ginseng* C. A. Meyer) is an important plant for medicine as well as for agriculture due to its unique pharmacological effect and high commercial value. Nevertheless, its genetic resource and study are still limited because of the difficulties in cultivation and handling of this plant. Especially, molecular marker for development and authentication of ginseng cultivars has not been actively studied. Therefore, in order to prepare basic resource for molecular marker, in this study, I investigated polymorphism present in *P. ginseng* cultivars and related *Panax* species, by AFLP and MITE analysis.

AFLP is an excellent technique for molecular marker development of plant species whose genome information is limited. Although genome sequence of *P. ginseng* is being actively analysed by using NGS technique, the generated draft sequence so far was imperfect and does not cover the whole genome of this species. Therefore, AFLP analysis was chosen for this study. First, 10 primer combinations showing high polymorphism were chosen after preliminary AFLP analysis between CP and YP and then used for 11 ginseng samples. As shown in Table 5, 117 polymorphic bands were identified between *P. ginseng* and *P. quinquefolium* and 5 bands among *P. ginseng* cultivars whose ratios were 21.3 % and 0.9%, respectively of the total 549 amplified bands. In a separate previous study using AFLP

technique previously published showed Polymorphic bands ratio of *P. ginseng* found in the wild (only in Russia) to be 2.9% [9].

This inefficiency in finding marker in my experiment may be due to low genetic variations, especially among ginseng cultivars. In fact, cultivars of *P. ginseng* have been pure-selected since 1970s [34] and thus genetic background among these cultivars is considered to be almost identical. On the other hand, *P. quinquefolium* was thought to be diverged from common ancestor with *P. ginseng* at 0.8 to 1.2 million years ago, based on analysis of synonymous substitutions per synonymous site (Ks) in paralogous genes [35]. Hence, genetic variation between *P. ginseng* and *P. quinquefolium* will be much higher than that among the cultivars. As a result, it is reasonable that more polymorphic bands were generated between the two ginseng species in this study.

Although AFLP analysis produced several polymorphic bands to be valued as molecular markers, its low efficiency and laborious work make it difficult to use it for further search of polymorphic DNA region. Therefore I introduced the other marker development systems based on MITE insertion polymorphism. MITE can be easily used for marker development because of its characteristics such as small size and high copy number, like SSR. In fact, MITE was used to develop marker in barley [36] and maize [37].

In this study, I identified 133 MITE consensus sequences in *P. ginseng* draft whole genome sequence database. Until now, MITE identification has been performed mainly in model plants whose whole genomes are available, such as Arabidopsis, rice, and Chinese cabbage. Therefore, the ginseng

MITE sequences found in this study is meaningful and also useful to understand genome structure of ginseng. Furthermore, considering low whole genome coverage of the draft sequence, ginseng genome will have more consensus sequences for MITE family.

As a final outcome, I analyzed 73 MITE regions of 25 MITE families and identified 16 polymorphic regions between *P. ginseng* and *P. quinquefolium*. The ratio of polymorphic region was 21.9% which was almost similar to that found in AFLP analysis. However, this similarity is thought to happen by chance, because all MITE families were not analyzed and the ratio of polymorphic MITE regions would be changed if complete whole genome sequence was used. In addition, 10 MITE regions were identified to show MIP pattern, of which 3 showed between *P. ginseng* and *P. quinquefolium* and rest 7 within single cultivar itself. MIP represents amplified length polymorphism depending on MITE presence and absence. The MIP was also reported in maize and barley [37]. All MIP bands found in *P. quinquefolium* were smaller than their counterpart bands in *P. ginseng*. This implies that MITEs causing MIP might be inserted in *P. ginseng* genome after divergence from *P. quinquefolium.*

All 10 MITE regions were identified to generate several amplicons showing MIP. Furthermore, many InDel and base substitutions were found among flanking sequences of MITE, when compared scaffold sequences matching to MIP amplicons. It indicates that *P. ginseng* genome was probably duplicated before insertion of the MITEs, although some MIP regions found in scaffolds seem to be resulted from MITE excision. To

estimate precise date of MITE insertion, further study with more MITE families will be needed.

Meanwhile, transposable elements have been considered to have an effect on regulation of gene expression. Recently, many studies have been published supporting that MITE play important roles in gene regulation [38-41]. Therefore, MITE in this study can be further studied using precise annotation of MITE location and whole ginseng genome sequences will be necessary.

In conclusion, this study analyzed genetic polymorphism in *P. ginseng* cultivars and related *Panax* species by AFLP and MITE analysis, and revealed polymorphic regions that can be used for molecular marker development. These results will be a valuable resource to understand structure and evolution of ginseng genome.

REFERENCES

1. Yun, T., Brief introduction of *Panax ginseng* CA Meyer. Journal of Korean Medical Science, 2001. 16(SUPP): pp. 3-5.
2. Shin, H.R., et al., The cancer-preventive potential of Panax ginseng: a review of human and experimental evidence. Cancer Causes and Control, 2000. 11(6): pp.565-576.
3. Attele AS, Wu JA, Yuan CS. Ginseng pharmacology: multiple constituents and multiple actions. Biochemical Pharmacology, 1999. 58: pp.1685-1693.
4. Jo, et al., Rapid identification of ginseng cultivars (*Panax ginseng* Meyer) using novel SNP-based probes. Journal of Ginseng Research, 2011. 35: pp.504-513.
5. Wang, et al., A PCR-based SNP marker for specific authentication of Korean ginseng (*Panax ginseng*) cultivar "Chunpoong". Molecular Biology Reports, 2010. 37: pp.1053-1057.
6. Lee, et al., CAPS markers using mitochondrial consensus primers for molecular identification of *Panax* species and Korean ginseng cultivars (*Panax ginseng* C. A. Meyer). Molecular Biology Reports, 2012. 39: pp.729-736.
7. Choi, et al., Development of reproducible EST-derived SSR markers and assessment of genetic diversity in *Panax ginseng* cultivars and related species. Journal of Ginseng Research, 2011. 35: pp.399-412.

8. Reunova, et al., Population genetic structure of wild-growing ginseng (*Panax ginseng* C.A. Meyer) assessed using AFLP markers. Russian Journal of Genetics, 2011. 48: pp.291-301.
9. Reunova, et al., Polymorphism of RAPD, ISSR and AFLP markers of the *Panax ginseng* C.A. Meyer (Araliaceae) genome. Russian Journal of Genetics, 2010. 46: pp.938-947.
10. Vos, P.R., Hogers, M., Bleeker, M., et al., AFLP: A new technique for DNA fingerprinting. Nucleic Acids Research, 1995. 23: pp.4407-4414.
11. Bureau, Wessler, Tourist: A large family of small inverted repeat elements frequently associated with maize genes. Plant Cell, 1992. 4: pp.1283-1294.
12. Wessler, et al., LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Current Opinion in Genetics and Development, 1995. 5: pp.814-821.
13. Jiang, et al., An active DNA transposon family in rice. Nature, 2003. 421: pp.163-167.
14. Yang, et al., Turned for transposition: molecular determinants underlying the hyperactivity of a *Stowaway* MITE. Science, 2009. 325: pp.1391-1394.
15. Paterson, Bowers, et al., The Sorghum bicolor genome and the diversification of grasses. Nature, 2009. 457: pp.551-556.
16. Oki, et al., A genome-wide view of miniature inverted-repeat transposable elements (MITEs) in rice, Oryza sativa ssp. japonica. Genes and Genetic Systems, 2008. 83: pp.321-329.

17. Smit, Riggs, Tiggers and DNA transposon fossils in the human genome. Proceedings of the National Academy of Sciences of the United States of America, 1996. 93: pp.1443-1448.
18. Bureau, Wessler, Mobile inverted-repeat elements of the tourist family are associated with the genes of many cereal grasses. Proceedings of the National Academy of Sciences of the United States of America, 1994. 91(4): pp.1411-1415.
19. Bureau, Wessler, Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell, 1994. 6(6): pp.907-916
20. Jiang, et al., Using rice to understand the origin and amplification of miniature inverted repeat transposable element (MITEs). Current Opinion in Plant Biology, 2004. 7(2): pp.115-119.
21. Amundsen, et al., Miniature inverted-repeat transposable element identification and genetic marker development in *Agrostis*. Crop Science, 2011. 51: pp. 854-861.
22. Santiago, et al., Genome-wide analysis of the Emigrant family of MITEs of *Arabidopsis* thaliana. Molecular Biology and Evolution, 2002. 19: pp.2285-2293.
23. Tu, Z., Eight novel families of miniature inverted repeat transposable elements in the African malaria mosquito, *Anopheles* gambiae. Proceedings of the National Academy of Sciences of the United States of America, 2001. 98: pp.1699-1704.

24. Chen, et al., MUST: a system for identification of miniature inverted-repeat transposable elements and applications to *Anabaena* variabilis and *Haloquadratum* walsbyi. Gene, 2009. 436: pp.1-7.
25. Han, Wessler, MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. Nucleic Acids Research, 2010. 38: e199.
26. Lu, et al., Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in Oryza sativa. Molecular Biology and Evolution, 2012. 29(3): pp.1005-1017.
27. Allen, et al., A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nature Protocol, 2006. 1: pp.2320-2325.
28. Vos, et al., AFLP: a new technique for DNA Fingerprinting. Nucleic Acids Research, 1995. 23: pp.4407-4414.
29. Rice, et al., EMBOSS: the European molecular biology open software suite. Trends Genetics, 2000. 16: pp.276-277.
30. Mathews, et al., Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. Journal of Molecular biology, 1999. 288: pp.911-940.
31. Zuker, M., Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research, 2003. 31: pp.3406-3415.
32. Bureau, et al., A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice

genes. Proceedings of the National Academy of Sciences of the United States of America, 1996. 93: pp.8524-8529.

33. Kuang, et al., Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs in the Solanaceae: new functional implications for MITEs. Genome Research, 2009. 19: pp.42-56.
34. Kang et al., An Illustrated Guide of Ginseng. Korea Agriculture Information Institute, 2010.
35. Choi et al., Genome duplication and speciation of *Panax ginseng* and *P. quinquefolius* inferred from comparative analysis of expressed sequence tags. BMC Genomics, 2012. Submitted.
36. Lyons, et al., Isolation, analysis and marker utility of novel miniature inverted repeat transposable elements from the barley genome. Molecular Genetics and Genomics, 2008. 280: pp.275-285.
37. Casa, et al., The MITE family Heartbreaker (Hbr): molecular markers in maize. Proceedings of the National Academy of Sciences of the United States of America, 2000. 97: pp.10083-10089.
38. Oki, et al., A genome-wide view of miniature inverted-repeat transposable elements (MITEs) in rice, *Oryza sativa* ssp. japonica. Genes and Genetic systems, 2008. 83: pp.321-329.
39. Amrani, et al., Genomewide distribution and potential regulatory functions of AtATE, a novel family of miniature inverted-repeat transposable elements in Arabidopsis thaliana. Molecular Genetics and Genomics, 2002. 267: pp.459-471.

40. Yang, et al., A two-edged role for the transposable element Kiddo in the rice ubiquitin2 promoter. The Plant Cell, 2005. 17: pp.1559-1568.
41. Naito, et al., Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. Nature, 2009. 461: pp.1130-1134.

ABSTRACT IN KOREAN

고려인삼(*Panax ginseng* C. A. Meyer)은 중요한 약리적 효능을 가진 오가과의 대표적 약용 작물이다. 고려인삼과 전칠삼(중국삼), 화기삼 (미국삼)은 전세계적으로 약용의 목적으로 이용되고 있으며 그 중에서도 고려인삼은 그 효능이 다른 종에 비하여 월등하여 높은 평가를 받고 있다. 고려인삼은 지금까지 9 품종이 개발되었는데 육성된 품종들 간의 식별 방법은 생육 과정 중에 경험적인 형태적 관찰을 통해서만 이루어지고 있기 때문에 보다 체계적이고 과학적인 방법이 요구되고 있는 실정이다. 따라서 본 연구의 목적은 고려인삼 품종들과 관련 종들을 구분할 수 있는 AFLP 와 MITE 를 기반한 분자 표지를 개발하는 것이다. AFLP 분석은 대상 식물의 유전체 정보가 없이도 분자 표지를 개발할 수 있는 기술이다. 본 연구에서는 AFLP 분석을 통해서, 증폭된 549 개의 밴드 중에 고려인삼과 미국삼을 구분하는 분자표지 117(21.3%)개를 찾았고 고려인삼 품종 간에 구분할 수 있는 표지 5(0.9%)개를 찾았다. 고려인삼 품종간 분자표지 개발의 효율성이 낮은 것은 품종간의 낮은 유전적 변이에서

기인한 것으로 생각된다. MITE 는 600 bp 이하의 작은 크기와 많은 복제수를 갖는 등의 특징으로 인해 보리나 옥수수 등 많은 식물체에서 분자표지를 찾는데 이용되어 왔다. 본 연구에서는 '고려인삼 유전체 시퀀싱 프로젝트' 에서 생산된 2.2 Gb 의 인삼유전체에서 133 개의 MITE consensus 서열을 추출하였고, 이중에 25 개의 MITE family 에 해당하는 73 개의 MITE 지역을 최종 분석하였다. 이로부터 고려인삼과 미국삼을 구분하는 16(21.9%)개의 분자표지를 찾았다. 그리고 73 개 MITE 지역중에 10 개가 MITE-based insertional polymorphism(MIP) 밴드 패턴을 보였는데, 이 중 3 개가 MITE 의 삽입 여부로 고려인삼과 미국삼을 구분할 수 있는 분자표지였다. 특히 MIP 중 MITE 삽입서열은 고려인삼에서만 나타나므로, 고려인삼과 미국삼의 종분화 이후 고려인삼에만 특이적으로 MITE 가 삽입된 것으로 추측되었다. 추가적으로, MIP 밴드 패턴을 보이는 10 개 서열을 분석하여 MITE 의 존재 유무를 확인하고 이들 서열간의 비교 분석을 수행하였다. 종합해보면, 본 연구에서는 AFLP 및 MITE 분석을 이용하여 고려인삼과 관련 인삼종의 유전적 다양성을 분석하였으며, 이로부터 향후 분자표지

개발에 활용될 수 있는 다양성 부위를 확인하였다. 본 결과는 인삼유전체의 구조 및 진화 연구에 유용한 자료로 쓰일 수 있을 것이다.