Research Article
J. Ginseng Res. Vol. 35, No. 4, 399-412 (2011)
http://dx.doi.org/10.5142/jgr.2011.35.4.399

JGR Journal of Ginseng Research

Development of Reproducible EST-derived SSR Markers and Assessment of Genetic Diversity in *Panax ginseng* Cultivars and Related Species

Hong-Il Choi[1], Nam Hoon Kim[1], Jun Ha Kim[1], Beom Soon Choi[2], In-Ok Ahn[3], Joon-Soo Lee[3], and Tae-Jin Yang[1*]

[1]Department of Plant Science, Research Institute for Agriculture and Life Sciences, Seoul National University College of Agriculture and Life Sciences, Seoul 151-921, Korea
[2]National Instrumentation Center for Environmental Management, Seoul National University College of Agriculture and Life Sciences, Seoul 151-742, Korea
[3]Natural Resources Research Institute, R&D Headquarters, Korea Ginseng Corporation, Daejeon 305-345, Korea

Little is known about the genetics or genomics of *Panax ginseng*. In this study, we developed 70 expressed sequence tag-derived polymorphic simple sequence repeat markers by trials of 140 primer pairs. All of the 70 markers showed reproducible polymorphism among four *Panax* species and 19 of them were polymorphic in six *P. ginseng* cultivars. These markers segregated 1:2:1 manner of Mendelian inheritance in an F_2 population of a cross between two *P. ginseng* cultivars, 'Yunpoong' and 'Chunpoong', indicating that these are reproducible and inheritable mappable markers. A phylogenetic analysis using the genotype data showed three distinctive groups: a *P. ginseng*-*P. japonicus* clade, *P. notoginseng* and *P. quinquefolius*, with similarity coefficients of 0.70. *P. japonicus* was intermingled with *P. ginseng* cultivars, indicating that both species have similar genetic backgrounds. *P. ginseng* cultivars were subdivided into three minor groups: an independent cultivar 'Chunpoong', a subgroup with three accessions including two cultivars, 'Gumpoong' and 'Yunpoong' and one landrace 'Hwangsook' and another subgroup with two accessions including one cultivar, 'Gopoong' and one landrace 'Jakyung'. Each primer pair produced 1 to 4 bands, indicating that the ginseng genome has a highly replicated paleopolyploid genome structure.

Keywords: *Panax* species, Expressed sequence tag-simple sequence repeat, Ginseng cultivars, Genetic diversity, Cultivar authentication

INTRODUCTION

Korean ginseng (*Panax ginseng* Meyer) is an important medicinal herb belonging to the family Araliaceae. Ginseng has been used as oriental medicine for thousands of years [1]. The major components showing pharmacological effects are the ginsenosides, which are known for their beneficial properties to the central nervous system, cardiovascular, endocrine and immune systems [2].

In ginseng research, medicinal components and their functions have been widely investigated. However, breeding, genetic and genomic studies have been rarely performed because of difficulty in maintaining plants and reproducing progenies. Approximately three to four years of growth is necessary to produce a small number of seeds, approximately 40 seeds per plant [3], thus hindering systematic management of genetic materials. Up to now, eight elite cultivars, 'Chunpoong', 'Yunpoong',

Received 15 Feb. 2011, **Revised** 08 Jun. 2011, **Accepted** 29 Jul. 2011

***Corresponding author**
E-mail: tjyang@snu.ac.kr
Tel: +82-2-880-4547, Fax: +82-2-877-4550

'Gumpoong', 'Gopoong', 'Sunpoong', 'Sunwon', 'Sunone' and 'Chungsun', have been bred by pure line selection and have been registered as commercial varieties since 1997 in Korea [4]. Even though the new varieties show better yields and qualities compared to that of the local landrace which is a native mixed line [3,4], they are being cultivated in less than 10% of the total ginseng cultivation fields because of the lack of well-organized seed production and supplying system. Stable seed supply systems with credible authentication method for each ginseng cultivar will promote high quality ginseng production via improvement of the ginseng breeding and seed industry.

Molecular breeding tools using DNA markers may be an alternative and indispensable way for ginseng improvement because marker assisted selection can reduce efforts and time for breeding. However, very limited numbers of DNA markers were also reported for ginseng. Random DNA markers such as random amplified polymorphic DNA [5-7] and amplified fragment length polymorphism (AFLP) [8] are used to study the diversities of local ginseng collections. However, these random primer-based markers could not be shared by common. Approximately, 60 of simple sequence repeat (SSR) markers have been produced from SSR-enriched libraries [9,10] and from bacterial artificial chromosome (BAC) end sequences [11] and are studied to determine the genetic diversity of ginseng collections. All of these SSR markers were derived from genomic sequences and these were not intensively studied between ginseng cultivars. Even though several papers described ginseng expressed sequence tags (ESTs) [12-15], there have been no reports on development of EST-derived SSR markers and their utilization in ginseng.

ESTs providing comprehensive transcript information [16] are valuable resources for development of molecular markers because they are derived from relatively conserved genic regions. EST-derived SSRs are also more advantageous than genomic SSRs because of the rich public availability of EST sequences and their high transferability to related species [17]. Thus, we are going to develop large number of EST sequence-derived SSR markers and construct a high resolution genetic map which can be utilized as a frame for genome sequencing. In this study, we tried to develop reproducible EST-derived SSR markers which can be applied for mapping and assessing a genetic similarity between registered commercial inbred varieties. And we estimated a polyploidy level of ginseng genome based on numbers of gene-based polymerase chain reaction (PCR) products.

MATERIALS AND METHODS

Plant materials and DNA extraction

Six *P. ginseng* accessions, four registered cultivars, 'Chunpoong', 'Yunpoong', 'Gumpoong' and 'Gopoong', bred by inbred line selection in Korea Ginseng Corporation (KGC) Natural Resources Research Institute (Daejeon, Korea), and two representative local landraces, 'Jakyung', mixed lines with red fruits, and 'Hwangsook', mixed lines with yellow fruits, and three related *Panax* species, *P. quinquefolius* originated in the USA, *P. japonicus* originated in Japan, and *P. notoginseng* originated in China, were included in the determination of genetic diversity. DNA pools derived from more than 15 individuals were used to represent each cultivar and landrace of *P. ginseng*. The DNA pool consisted of a mixture of the same amount of template DNA from 15 individuals of each cultivar and landrace of *P. ginseng* and from five individuals of *P. quinquefolius*. However, single individual DNA was used to represent *P. japonicus* and *P. notoginseng* because of limited materials. An F_2 population that consisted of 51 individuals from a cross between 'Yunpoong' and 'Chunpoong' was used to determine the inheritability and reproducibility of the newly developed markers. All leaf samples were kindly provided from KGC Central Research Institute. Total DNA was extracted using the modified cetyltrimethylammonium bromide method [18]. DNA concentrations were measured using ND-1000 (NanoDrop Technologies Inc., Wilmington, DE, USA).

Construction of the ginseng expressed sequence tag database and repeat motif screening

A ginseng EST database was constructed by collecting sequences from public databases. After removal of poly-A tails using PanGEA [19], repeat-oriented sequences and SSR motif-containing sequences were characterized from the raw EST database using RepeatMasker ver. 3.2.6 (http://www.repeatmasker.org) which was downloaded and installed on the local computer. In the screening process, a default mode with a "-poly" option was used to select genuine SSR motifs.

Designing primers

We extracted ESTs which contained 3-6 copies of SSR motifs using Tandem Repeat Finder [20]. Primer pairs were designed from the flanking sequences of SSR motifs (Table 1) with 18 to 27 bp nucleotides using the Primer3 program (http://frodo.wi.mit.edu/primer3/). Product sizes ranged from 150 to 600 bp. Standalone BLAST executables (BLASTN 2.2.15, ftp://ncbi.nlm.

Table 1. Simple sequence repeats found in ginseng EST sequences and summarization of SSR marker developments

Repeat unit (length)	SSR[1)]	Primer design[2)]	PCR Success (%)	Polymorphic between	
				Panax ginseng cultivars (%)	*Panax* species (%)
Mono	855	11	4 (36.4)	0	0
Di	379	28	23 (82.1)	9 (39.1)	16 (69.6)
Tri	246	45	41 (91.1)	9 (22.0)	30 (73.2)
Tetra	87	10	8 (80.0)	0	2 (25.0)
Penta	70	12	12 (100)	1 (8.3)	11 (91.7)
Hexa	34	5	4 (80.0)	0	3 (75.0)
Degenerate	-	29	27 (93.1)	0	8 (29.6)
Total	1,671	140	119 (85.0)	19 (16.0*)	70 (58.8)

EST, expressed sequence tag; SSR, simple sequence repeat; PCR, polymerase chain reaction.
[1)]No. of potentially polymorphic SSRs based on the poly option of the RepeatMasker program.
[2)] No. of sites used for primer designing.

nih.gov/blast/executables) were used to avoid primer design in duplicate sequences. Various sources of SSR markers reported in the previous papers were designed and tested together [9,10].

Polymerase chain reaction and electrophoresis

PCR amplifications were performed in a 25 μL volume with 1 U *Taq* DNA polymerase (Vivagen, Seongnam, Korea) according to the manufacturer's protocol using a DNA Engine Thermal Cycler (Bio-rad, Hercules, CA, USA). Conditions for the PCR cycle were as follows: 5 min at 95°C for denaturation, 38 cycles of 10 s at 95°C, 30 s at T_m°C, 20 s at 72°C, and 10 min at 72°C for final extension. PCR products were separated on 2% agarose gels and on 5% denaturing polyacrylamide gels or 9% non-denaturing polyacrylamide gels.

Data analysis

Blast2GO ver. 2.4.2 was used to annotate polymorphic ESTs with default parameters [21]. Analyses of developed marker data were conducted using PowerMarker ver. 3.25 software [22]. Allele frequency data were obtained as a binary matrix and were imported into NTSYSpc 2.11X (Exeter Software; Setauket, NY, USA) for phylogenetic analysis using Dice's coefficient [23] and the unweighted pair group method with the arithmetic mean [24]. Bootstrapping of the tree with 1,000 replications was generated in Winboot [25].

RESULTS AND DISCUSSION

Simple sequence repeat motif in ginseng expressed sequence tag sequences

A total of ca. 11 Mb, 19,578 ginseng ESTs, were collected from the public database and reconstructed in our local server (http://im-crop.snu.ac.kr). Using an EST trimming process, ca. 17.5 kb of poly-A tails were removed and 1.6% (179,540 bp) of ESTs were masked as repeat sequences. Of them, 1,584 bp showed a significant homology with 19 kinds of transposable elements, 8,619 bp had a significant homology with 32 non-coding RNA elements, and 62,348 bp derived from 1,344 regions were screened as low complexity DNA. A total of 2,158 regions spanning 108,116 bp were classified as simple sequence repeats, and 1,671 sites spanning 68,763 bp detected in 1,567 ESTs (8.0% of the total raw ESTs) were classified as potential polymorphic SSR sites by the "-poly" option (Table 1). Most of the ESTs (94.6%) contained one SSR motif, with exceptions containing two SSRs (4.6%) or three to five SSRs (0.8%).

Classification based on the nucleotide lengths of the repeat motifs showed that 51% of SSRs were composed of mono-nucleotide motifs, but mononucleotide repeat motifs showed poor success rates for PCR as well as polymorphism that may have derived from "stutter" artifacts in the PCR reaction [26]. When we ignored mononucleotide repeats, di- and tri-nucleotide repeats were the most abundant classes, including 46.4% and 30.1%, respectively. Similarly, di- nucleotide repeat motifs were also predominant in kiwifruit, spruce and coffee [27-29]. Meanwhile, tri-repeat motifs were most predominant in the EST-SSRs of grape, sugarcane, barley, wheat, maize, sorghum, rice, rye, oats, and flax [30-36].

Assessment of homogeneities in ginseng landraces and registered cultivars

During development of DNA markers polymorphic between accessions, we found that DNA markers were

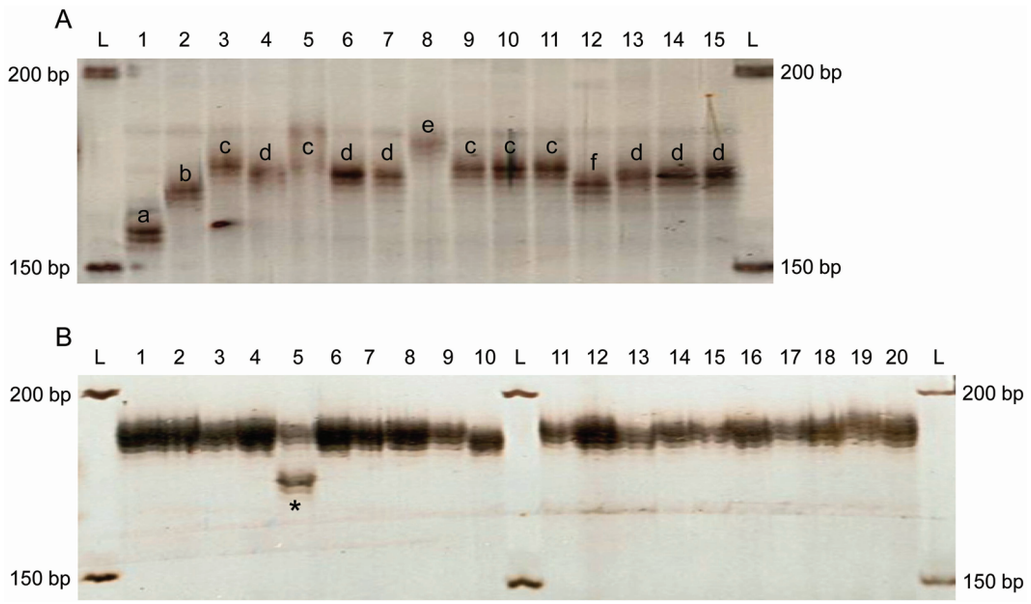

Fig. 1. Allelic variations among individuals of a landrace and a cultivar. Denaturing polyacrylamide gel electrophoresis was conducted for separation of PCR products using individual DNA and the GES0002 marker. (A) A total of 15 individuals of landrace 'Jakyung' were surveyed. Different genotypes were denoted as a-f and genotype d is shown as major. (B) A total of 20 individuals of cultivar 'Chunpoong' were surveyed. Only number 4 plant denoted by * shows heterozygous allele. L, DNA ladder; GES, ginseng expressed sequence tag-simple sequence repeat.

not homogeneous among individuals in one local landrace accession or in the registered cultivars, as shown in Fig. 1. The most popular landrace 'Jakyung' individuals showed up to six different genotypes among 15 individuals, even though all were derived from the same seed lot (Fig. 1A). Meanwhile, one of the elite cultivars, 'Chunpoong', showed a relatively uniform genotype with two off-types among 20 individuals randomly selected from the same seed lot (Fig. 1B). Screening with three markers developed in this study showed a relative range of heterogeneity of 10% to 30% in six accessions. Also, we found that previously reported polymorphic ginseng markers were not reproducible in our trial [9,10,31]. This may be due to the heterogeneity of the ginseng population because they used an individual plant DNA for representative of each accession.

Approximately 56% of AFLP bands showed polymorphism among wild *P. ginseng* individuals in Russian Primorye area and population structure study clearly differentiated their phylogenetic relationships based on frequencies of individual alleles [37]. Similarly, more divergence detected in wild *P. quinquefolius* than the cultivated [38] that might be derived by possible out-crossing events even though the plants prefer self-fertilization [39]. And cultivated *P. notoginseng* population remained fair level of biodiversity, ranged 74% to 39% of divergence depends on location in China [40]. Our result showed abundant genetic diversity even in cultivating *P. ginseng* landraces (Fig. 1A) that might be derived from bulked seed harvesting from genetically unfixed lines and also from temporal out-crossing [39]. Meanwhile, eight elite cultivars were bred by pure line selection even though each showed approximately 10% of off-type allele. Therefore, we considered that utilization of a DNA pool from many individuals will be credible to identify more reproducible markers than using a single plant for representing each Korean ginseng cultivar even though the method can ignore many rare alleles. We concluded to use a DNA pool derived from 15 individual plants to represent each Korean ginseng cultivar because our purpose is development of markers which can be a general representative for each elite cultivar (Fig. 1B).

Development of simple sequence repeat markers and their transferability to related species

We designed a total of 140 primer pairs amplifying 111 SSR motifs and 29 degenerated SSR motifs (Table 1). Among them, 119 pairs produced bands of which 105 were the same as the expected sizes and 16 were larger than the expected. And 21 primers produced no clear PCR product. The PCR failure rate was 15% for the 140 trials of EST-SSR primers, which is in the 10% to 40% range reported in previous EST-SSR analyses for other species [27,30,31,35,41-44]. Seventy pairs, correspond-

ing to 58.8% of the successful PCR primers, showed polymorphism for at least one of nine accessions including six *P. ginseng* accessions and three related species. Polymorphisms were mainly restricted to interspecies, and only 19 of 70 SSRs showed polymorphism between *P. ginseng* cultivars as well as between *Panax* species and were named ginseng EST-SSR (GES) (Fig. 2A, B; Tables 1 and 2). The other 51 markers were polymorphic only among *Panax* species. Among those, 43 were derived from intact SSR regions and were named *Panax* EST-SSR (Fig. 2C, D; Tables 1 and 2). Eight were derived from degenerated SSR regions and were named *Panax* EST (Tables 1 and 2). BLASTX analysis revealed that 51 of 70 polymorphic ESTs showed significant hits and 19 did not have any matches with the known proteins. Among the significant 51 hits, 22 showed best hits with genes in *Vitis vinifera* (Table 2).

The mean level of polymorphism was 20% at the species level that is similar to those of previous EST-SSR studies in other species [30,35,45-48]. The level of polymorphism in EST-SSR is lower than that of genomic SSR primers because the transcribed regions are more conserved than the non-coding regions [17,29,34,49-52]. Our gene-based SSR markers showed comparable levels of polymorphisms to those of the genomic SSR markers for distinguishing *P. ginseng* accessions. Only 22 of

Table 2. Characteristics of 70 polymorphic expressed sequence tag-simple sequence repeat loci in *Panax ginseng* cultivars and related species

Marker	Primer pair (5' → 3')	Repeat motif	T_a (°C)	Size[1]	N_r[2]	N_a[3]	MAF	GD	PIC	N_b[4]	Sequence description	Min. E-value
GES0001	GCATGGCAATTTGGAGAGAGGTACG	(GAA)n	60	196	14	7	0.3333	0.8148	0.7938	3	No hits found	
	GTTCTGTGACTTGCCGGTTTGCTCC											
GES0002	GTGGTGAGAAAGGGAAAGAGCAATCG	(TA)n	60	176	22	7	0.2222	0.8395	0.8194	2	No hits found	
	CCCTCGATCTACAGATGATCAAATAGC											
GES0003	TTTCAAATGGATCTATGAGAATAATGA	(TTC)n	54	247	20	5	0.4444	0.7160	0.6773	3	No hits found	
	TGGGCACATAAAAAGACAGTG											
GES0004	CTAGTCAACAACATCATCATCATCC	(TA)n	56	213	27	5	0.3333	0.7407	0.6987	4	Sugar isomerase	1.E-12
	TGCAGAATTAACTGAGACTCAAGAA											
GES0005	TTCTTCCTTGCACGTTTCTACTACT	(CCA)n	56	190	14	7	0.2222	0.8395	0.8194	2	at5g14920 f2g14_40	2.E-28
	AATATAATTGCTACACTCCCCTTGG											
GES0006	AGCCTAGTGTGCAGAAGTAAAGTGT	(TA)n	56	238	26	4	0.4444	0.6667	0.6072	3	Protein	2.E-15
	TGAAGTAGAACTGATCACAGAGTGC											
GES0007	GGGGCTTCTCTAATTTACACCTTTA	(GA)n	55	243	16	4	0.4444	0.6914	0.6401	3	Protein	9.E-41
	AAAGATGAAAACTTGATGCTTGTTC											
GES0008	TGGTCTAGAACAGAAAAGATCGAGT	(TA)n	55	187	28	5	0.4444	0.7160	0.6773	2	No hits found	
	GTACTGTCTGTGGTTTGAATGATTG											
GES0009	TGAACCACATGATTTACGATTAGTG	(TA)n	56	240	11	6	0.3333	0.7901	0.7615	2	Xyloglucan endotransglycosylase	9.E-13
	GACATATCTGCATGGCTTTCTTAAT											
GES0010	AGGACTTCAATGCTAGAACTCAGAA	(TA)n	56	285	17	4	0.4444	0.6914	0.6401	2	Mitochondrial ribosomal protein l11	2.E-44
	CATGGGCTAAATAATAAAAGACCAA											
GES0011	GTTATGACCGGTAAATTAGGTTGGT	(TA)n	56	229	11	4	0.5556	0.6173	0.5688	2	Protein	2.E-46
	CAATCCACATCAAGACCATATTACA											
GES0012	TAATTATATTTGTGTTGGCAGACGA	(TC)n	54	227	15	5	0.4444	0.7160	0.6773	3	Predicted: hypothetical protein [*Vitis vinifera*]	4.E-13
	CTCGGCATACAACATTTAACTTACC											
GES0013	ATTAGTAGAACGTACAGCCCAAACC	(CCA)n	56	230	8	5	0.4444	0.7160	0.6773	2	Carboxyphosphonoenolpyruvate mutase	5.E-58
	TATGGTAACTTTAGGCTGGTGTAGC											
GES0014	GAGAAAATTGAGGAACCAAACAAG	(CTA)n	54	299	12	4	0.6667	0.5185	0.4847	4	No hits found	
	GTTTTCTCCACAACTACTGGCTCT											

Table 2. (Continued)

Marker	Primer pair (5' → 3')	Repeat motif	T_a (°C)	Size[1]	N_r[2]	N_a[3]	MAF	GD	PIC	N_b[4]	Sequence description	Min. E-value
GES0015	AAAATTCTGCTCACACTCTCTCTGT	(CTA)n	56	193	10	4	0.5556	0.6173	0.5688	2	No hits found	
	CGGAGTTTTTGAAGATAAGAATCAA											
GES0016	ATTTATATATCTTCACGCTGCTTCG	(TCC)n	56	230	10	4	0.4444	0.6667	0.6072	3	Beta-galactosidase a-peptide	5.E-13
	CAAAAATAAGAGATGGAGATGGAGA											
GES0017	AAAATGGTTCCAAATTGTGCTTC	(TTA)n	56	239	11	5	0.3333	0.7654	0.7279	4	No hits found	
	AAGGTGGAAATAAGGAGAGAAAAGA											
GES0018	CTCTCTCTCTCTCTCTCTCATCTGC	(TTC)n	56	170	8	4	0.6667	0.5185	0.4847	2	Protein	1.E-46
	AAAGAAGAACCACAAACACTAAACG											
GES0019	GTACTATGGATAAAGCTGGAATGGA	(TAGGG)n	56	207	6	5	0.3333	0.7654	0.7279	2	No hits found	
	CGGTAAGTGACACTAAGAACAACTG											
PES0001	GGAGCAGCAATAGACCAAGG	(CCCTG)n	55	356	4	2	0.7778	0.3457	0.2859	3	Predicted: hypothetical protein [*V. vinifera*]	1.E-04
	TTGTTTGGAAACCTGGGAAC											
PES0002	TCGGAGCGAGAAGAAGAGTC	(ATG)n	56	290	9	3	0.7778	0.3704	0.3402	2	Nascent polypeptide associated complex alpha chain	1.E-68
	CGTCTTCATCATCCTGAGCA											
PES0003	GGTGGAGATCACAAGGAAGG	(GAA)n	56	312	8	3	0.7778	0.3704	0.3402	2	No hits found	
	TGGCAACAATCAGCATCCTA											
PES0004	CGAAGGTGCACCAAAAGTCT	(CACCAT)n	56	365	5	2	0.7778	0.3457	0.2859	4	No hits found	
	GGACGAAGACGTGGCTCTAC											
PES0005	TGGGTTCAACTTTGGAGGAG	(CAGGT)n	56	243	11	3	0.7778	0.3704	0.3402	2	Protein	5.E-15
	CTCTTTCACCGCAACAGACA											
PES0006	CAACCTTTTAATTCCTTTGCTACA	(CAT)n	54	172	10	2	0.7778	0.3457	0.2859	2	Transcription factor gt-3a	1.E-08
	CCGTCTCAATATTCACACTGATCT											
PES0007	CGAGGAGTCAAAGGTGGAAG	(GAA)n	56	266	15	2	0.7778	0.3457	0.2859	2	Dehydrin	1.E-23
	CGCCTGGAAGTTTTTCTTTG											
PES0008	AACGTGATGCATGTCGAGAG	(TA)n	54	176	26	3	0.7778	0.3704	0.3402	3	Catalase	1.E-24
	GCACCGAGTTTTCCCAAGTA											
PES0009	GGAGGCCCGACTTACCTACT	(GGC)n	54	213	6	2	0.8889	0.1975	0.1780	1	No hits found	
	CACGTTGACGTGGCTATCTG											
PES0010	GTCTCGCAAAGAATGTCAGC	(CCA)n	55	189	7	2	0.7778	0.3457	0.2859	2	g1-like protein	2.E-47
	CTGCTTTTGCACCTCATAGC											
PES0011	TATCCACCAAAACACTACTCATCCT	(ATG)n	54	258	7	2	0.8889	0.1975	0.1780	1	Predicted: hypothetical protein [*V. vinifera*]	2.E-05
	CCTCTTAGACTCGTCATTAGGTTCA											
PES0012	ATTTAGCTTGGCTATATGTGAATGG	(CAG)n	54	284	6	3	0.7778	0.3704	0.3402	2	No hits found	
	GGACAGAAGTGAAGCATTTCATAGT											
PES0013	TCCTAAATTAGCACTAAACGCACAT	(CAG)n	54	162	8	3	0.7778	0.3704	0.3402	1	Dna binding	1.E-37
	TTGTTTACTAAATTCATGGGAGAGG											
PES0014	CAACTGCAAAGTCAAAATAATACGA	(TA)n	56	180	15	3	0.7778	0.3704	0.3402	2	Myb-like transcription factor 1	5.E-21
	GTAATCTTCCAGCTATCAAAGACCA											

Table 2. (Continued)

Marker	Primer pair (5' → 3')	Repeat motif	T_a (°C)	Size[1]	N_r[2]	N_a[3]	MAF	GD	PIC	N_b[4]	Sequence description	Min. E-value
PES0015	ACAAGAACAATTGTCAAAGGAAGTC	(TA)n	56	300	12	3	0.7778	0.3704	0.3402	1	Ribulose bisphosphate small subunit	1.E-62
	CTTTCAACACCTGAGATGAATCAGT											
PES0016	GAATGATCATATATACCTCCACTGGT	(TA)n	54	252	11	3	0.7778	0.3704	0.3402	2	Dehydrin 7	2.E-50
	TAAAATAAGCATTAGCAGAGCCATC											
PES0017	TCGGTAAGGATATCATCAACAAAAT	(TC)n	54	174	10	3	0.7778	0.3704	0.3402	2	Sterol carrier protein 2-like	2.E-18
	TTTTTGATAAAGACAAGGTCAAAGC											
PES0018	GGTATTGCTCGTGAACTTTGTAACT	(TC)n	54	201	20	3	0.7778	0.3704	0.3402	2	Aspartyl protease	8.E-107
	CAATAGGAAGAGAAGAAAACCAACA											
PES0019	GGAAACAGGGGTAGAAGAAGTGTAT	(TC)n	56	287	10	3	0.7778	0.3704	0.3402	4	Dcn1-like protein 4	2.E-70
	AGTATTTGTTGTTCCTTTCCTGGAT											
PES0020	CTATACCTCAGCACCAGTTTCAACA	(CAG)n	54	298	6	3	0.7778	0.3704	0.3402	2	S-rnase-binding protein	9.E-63
	TATCTGCGAATTATTTTCCATGAAT											
PES0021	GAAAACATTTGTGTTTCAGTAGGC	(CAG)n	56	155	10	3	0.7778	0.3704	0.3402	2	60s ribosomal protein	5.E-92
	AATGAGCTTCAGGTAAATATCATCG											
PES0022	CCAAGCCACATAAATCTAGGAGTATC	(CAG)n	56	154	7	3	0.7778	0.3704	0.3402	2	No hits found	
	AAAAACAACAAGTGCAGTTACACAA											
PES0023	CACAGTGAGGAAGAAGAAGAAGAAG	(CAT)n	56	152	8	2	0.7778	0.3457	0.2859	3	60s ribosomal protein l6	1.E-54
	ACCTGGAATACTTTCCAATACCG											
PES0024	TAATAATAATTTGATGCGGTTCCAT	(CAT)n	56	172	10	2	0.7778	0.3457	0.2859	2	Protein	5.E-24
	GGTGTTGTCAATTAGGAGAGAGAAA											
PES0025	AAAATCAATCTCCCATAAATTTGGT	(CTG)n	56	219	8	3	0.7778	0.3704	0.3402	2	Serine threonine-protein kinase	1.E-13
	TGATGTTTTGAAACAGAATCTTCAA											
PES0026	AGAATTTGAAGATGATGAAGAATCG	(GAA)n	56	214	9	3	0.7778	0.3704	0.3402	2	Protein	9.E-81
	AAAAGCTTTAGCCAAAGAAAGAGAG											
PES0027	ACTTTTATCCCAAAGCATCTTTTCT	(GAA)n	55	181	7	3	0.7778	0.3704	0.3402	1	No hits found	
	GTTCTACCTCTGAATTGGCACTAGA											
PES0028	GATACCTCAACAAAAATCCATCAAC	(GGA)n	55	154	9	3	0.7778	0.3704	0.3402	3	Protein	3.E-70
	ATGACGTTGTGCTTTTATAGCTTCT											
PES0029	TACTCCTTCCCCATTTAATTTCTTC	(TCC)n	56	183	8	3	0.7778	0.3704	0.3402	2	Ap2 erf domain-containing transcription factor	9.E-29
	GAGGGCAGTTTTGGTAGATATTTTT											
PES0030	AACCTAATGTCCGGAAGGTAGTAAC	(TTA)n	56	259	9	3	0.7778	0.3704	0.3402	3	Receptor protein	2.E-42
	CAATGTATGGAGGAAGGTTTATTTG											
PES0031	CGAATTCTGATTCTTGACATTTCTT	(TTC)n	55	210	7	3	0.7778	0.3704	0.3402	4	Protein aq_1857	7.E-28
	CGAAATATGGAGTAAGACGCTGTAT											
PES0032	GTGAAGCTTGAACACTTAGAAGAGG	(TCCC)n	55	165	6	3	0.7778	0.3704	0.3402	3	Oligosaccharyl transferase	3.E-123
	TCATCATACTTTGCTAACACGTCAC											
PES0033	AAAACGAAGAAAATGTTACAAGTGC	(TCTA)n	56	298	6	3	0.7778	0.3704	0.3402	2	No hits found	
	GTATCACCACAACATCAAATTACCA											

Table 2. (Continued)

Marker	Primer pair (5' → 3')	Repeat motif	T_a (°C)	Size[1]	N_r[2]	N_a[3]	MAF	GD	PIC	N_b[4]	Sequence description	Min. E-value
PES0034	GCTTATCTTGCCTGATAATGTCCTA	(ATCTG)n	56	259	5	3	0.7778	0.3704	0.3402	2	Protein	1.E-59
	CAACATGTAATTACGTCTCTCATGC											
PES0035	CAATTTCTTCCGTATAAACCAATTT	(AATCA)n	56		5	2	0.7778	0.3457	0.2859	1	Predicted protein [*Populus trichocarpa*]	5.E-06
	ACTCCTACCTGCACAATTTGAATAC											
PES0036	AGTGGCAACTCTAGAGAAACTATGC	(CACCC)n	55	259	5	3	0.7778	0.3704	0.3402	1	Conserved hypothetical protein [*Ricinus communis*]	2.E-22
	ATACGTTACATGGCAGCTGAATACT											
PES0037	CAAAAGGCTTGCTCTATATTGTGAT	(CAGGC)n	56	275	4	3	0.7778	0.3704	0.3402	2	Ras-related small gtp-binding protein	1.E-93
	TCCAGTCATTAATACTTGCAACAAA											
PES0038	TCTCCTTCTGCTGGTAGTAAAAATG	(CCGTA)n	56	190	4	3	0.7778	0.3704	0.3402	2	Unnamed protein product [*Vitis vinifera*]	5.E-38
	ATAAATTCTGCATCTACGGTGTAGC											
PES0039	TGATCCAGTTCAATTTCTTGTTTTC	(GTTTG)n	56	170	4	3	0.7778	0.3704	0.3402	2	No hits found	
	ATAATCTTTCAATTTCCCGTACACA											
PES0040	ATTGTTGAAGAAATTGGTTGTTTGT	(TTTTC)n	56	270	6	3	0.7778	0.3704	0.3402	3	Phosphate phosphoenolpyruvate translocator precursor	1.E-36
	CTGAGTTCATTTCCTGGAACATAGT											
PES0041	ACTGTTTATGACTGGCTCTACAGTG	(TTTTG)n	56	282	9	3	0.7778	0.3704	0.3402	1	Auxin-repressed protein	2.E-29
	GTACCGTCCATGACATAACATAACA											
PES0042	TAAATTCACTTTGTGTGTGTGTGTG	(GGGAGA)n	56	229	5	3	0.7778	0.3704	0.3402	2	O-methyltransferase-like protein	6.E-89
	GTTCTTGATCGTGATTCTTTTCAAG											
PES0043	GGGAGACTAATTTTCTTTGCTTTTC	(TTCTCC)n	56	224	3	3	0.7778	0.3704	0.3402	2	F-box family protein	2.E-29
	TTCTGATGAGTTCTGTGTAGGAGTG											
PE0001	GCCCTAGCCCTAATCAATCC	-	55	314	-	2	0.7778	0.3457	0.2859	3	At4g30930-like protein	9.E-25
	GGGCCAATGACCTTATACCC											
PE0002	GATCTCGAACCGACGAACTC	-	54	376	-	3	0.6667	0.4938	0.4377	1	No hits found	
	AACCATACTGCCAACAATTAAGC											
PE0003	GCCTTCTGAACTTCCTGGTG	-	54	154	-	2	0.7778	0.3457	0.2859	1	Eukaryotic translation initiation factor 3	3.E-11
	GTCAGGTTCTGCAGGTGGA											
PE0004	GGTTCTCGGGACAATGAAAG	-	54	219	-	2	0.7778	0.3457	0.2859	2	Unnamed protein product [*Vitis vinifera*]	6.E-09
	ACCCCATTCCCTTCTCTCAC											
PE0005	GCAACATCACCGTCAATGAG	-	56	264	-	3	0.6667	0.4938	0.4377	1	Glycine-rich rna-binding protein	2.E-37
	CACAAATTTACCAGCCACCA											
PE0006	TCCTCTGCCACATTTAAGCA	-	55	309	-	2	0.7778	0.3457	0.2859	2	No hits found	
	TCATGTTGCAAGAGCAAAGC											
PE0007	GTGGAAGAGGCAAAACCAAG	-	56	210	-	3	0.7778	0.3704	0.3402	2	Nucleic acid binding	8.E-44
	AGCCATGCTAGGTCTGTTGG											
PE0008	GGTCTTGGTCTTGGAGTTGG	-	54	221	-	2	0.6667	0.4444	0.3457	1	No hits found	
	CCTCCTTGATTTCCACCTGA											
Average						3.3143	0.6810	0.4568	0.4176	2.2		

MAF, major allele frequency; GD, genetic diversity; PIC, polymorphic information content; GES, ginseng expressed sequence tag-simple sequence repeat; PES, *Panax* expressed sequence tag-simple sequence repeat; PE, *Panax* expressed sequence tag.
[1]Expected amplicon size; [2]No. of repeats; [3]No. of alleles; [4]No. of bands around the expected size in *P. ginseng*.

189 (representing 11.6%) [10] and 11 of 94 genomic SSR primers (11.7%) [9] were polymorphic among *P. ginseng* accessions when the primers were designed using a microsatellite-enriched library. Meanwhile, 12 out

of 31 BAC end sequence-derived genomic SSR primers (38.7%) were polymorphic among *P. ginseng* accessions [11,53].

Even though not all of the nucleotide repeat units were surveyed by PCR, our trials showed that big differences in the rates of successful PCR and the appearances of polymorphisms depended on the repeat unit length. SSRs with penta- and tri-nucleotide repeat motifs showed the highest degrees of PCR success and polymorphism detection between *Panax* species. SSRs having a di-nucleotide motif showed the highest polymorphism rates among the ginseng cultivars. Meanwhile, SSRs derived from mono-nucleotide polymers were not optimal for PCR amplification or polymorphism detection (Table 1).

Transferability among related species is considered the most important feature for EST-SSR markers that help to produce conserved orthologous markers and thus be applicable to related species which have little genomic information [17]. In this study, primers designed from *P. ginseng* ESTs were successfully used in the related species of *P. japonicus*, *P. quinquefolius* and *P. notoginseng* with 100%, 97.1%, and 75.7% transferabilities, respectively, that is similar to the previous studies shown 100% transferability between *P. ginseng* and *P. quinquefolius* [9,11].

Number of bands and estimation of the polyploidy level in *Panax ginseng*

Recent progress in the field of genomics has uncovered highly replicated polyploidy levels in most of the plant genome [54,55]. The *P. ginseng* genome is considered as tetraploid because of the chromosome number variations, 12 *vs*. 24 pairs [56]. However, there has been no molecular evidence to determine their ploidy level or to identify a polyploidization event in the *Panax* species. Polyploidy levels were previously studied in the olive complex (*Olea europaea*) based on the band numbers of highly polymorphic SSR markers [57]. Various allele numbers were detected in various subspecies, and maximums of four and six alleles were detected in tetra- and hexaploid subspecies, respectively, that were consistent with those of the flow cytometry analyses.

To estimate copy numbers of homologous genes and thus assume the polyploidy level, we counted the number of bands around the expected size in four relatively homogeneous inbred cultivars by assuming that different bands may have been derived from recently duplicated paralogous genes. Electrophoresis of the PCR products revealed various band patterns around the expected sizes. Out of the 119 successful SSR primer pairs, only 17 pairs yielded one specific target band, as shown in Fig. 2C, and 49, 22, and 10 pairs produced two, three, and four bands, respectively (Table 2 and Fig. 2). The other 21 pairs yielded unspecific faint bands. Overall, the data indicate that over 85% of the genes remained as duplicate genes with one to three extra paralogous gene copies, thus indicating that ginseng has a highly replicated polyploidy level which may range from a tetra- to octa-paleoploidy genome. These results are similar to or greater than the polyploidy level

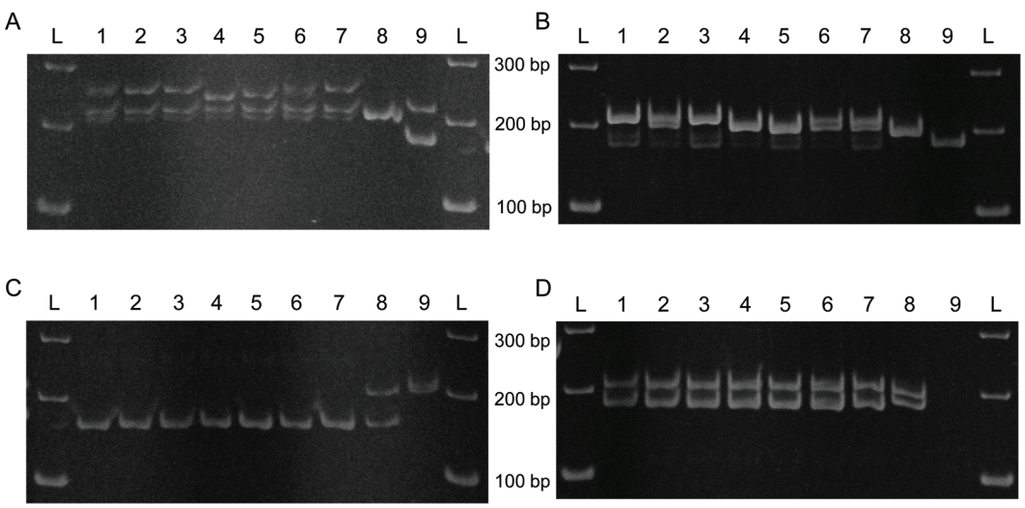

Fig. 2. Non-denaturing polyacrylamide gel electrophoresis of PCR products from nine *Panax* accessions using different markers, (A) GES0003, (B) GES0019, (C) PES0010 and (D) PES0034. Lanes 1-6 indicate *P. ginseng* accessions, Gumpoong (1), Hwangsook (2), Yunpoong (3), Chunpoong (4), Gopoong (5), Jakyung (6) and related *Panax* species, *P. japonicus* (7), *P. quinquefolius* (8), *P. notoginseng* (9). L, DNA ladder; GES, ginseng expressed sequence tag-simple sequence repeat; PES, *Panax* expressed sequence tag-simple sequence repeat.

suggested to be a natural tetraploid [56,58].

Different band numbers were detected in two landrace accessions because of their heterogeneity among individuals, as shown in Fig. 1A. One major polymorphic band was detected in four ginseng cultivars, but two clear bands were observed in two landraces, Hwangsook (2) and Jakyung (6), presumed to be derived from different alleles in two groups of individuals in the landraces because we used a DNA pool derived from 15 individual plants (Fig. 2B). Differences in band numbers were also detected in different species, such as in lane 7 in Fig. 2B and lane 8 in Fig. 2C, that may have been derived from a difference in gene copy numbers. One clear band was amplified in most accessions, but two bands were produced in a single *P. japonicus* plant and a *P. quinquefolius* DNA pool derived from five individuals, indicating that the species are heterozygous allele or included an extra paralogous gene.

Phylogenetic analysis of ginseng cultivars and related species

Because most amplicons showed multi-band profiles, genotyping was limited only to the major bands which appeared around the expected size. Variations in amplicon sizes were manipulated as unweighted and independent characteristics. Major allele frequencies were in the range of 0.2222 to 0.8889, with an average of 0.6810. The number of alleles was in the range of two to seven, with an average of 3.3143. Gene diversity and polymorphism information content ranged from 0.1975 to 0.8395 (average, 0.4568) and 0.1780 to 0.8194 (average, 0.4176), respectively (Table 2).

A phylogenetic analysis of the nine accessions was conducted using 215 allelic data points produced from 70 markers. Three clades were separated at similarity coefficients of 0.7: *P. ginseng*-*P. japonicus* clades, *P. notoginseng* and *P. quinquefolius* (Table 3 and Fig. 3). It is

Table 3. Dice's similarity coefficient matrix for nine accessions obtained from 70 marker data

	Chunpoong[1)]	Yunpoong[1)]	Gumpoong[1)]	Gopoong[1)]	Hwangsook[1)]	Jakyung[1)]	*P. japonicus*	*P. quinquefolius*	*P. notoginseng*
Chunpoong*	1.0000								
Yunpoong*	0.7571	1.0000							
Geumpoong*	0.7571	0.8429	1.0000						
Gopoong*	0.8000	0.8286	0.8000	1.0000					
Hwangsookjong*	0.7857	0.8571	0.8429	0.8000	1.0000				
Jakyungjong*	0.8143	0.8714	0.7857	0.8714	0.8571	1.0000			
P. japonicus	0.7286	0.8143	0.8286	0.7857	0.8000	0.7857	1.0000		
P. quinquefolium	0.0286	0.0143	0.0143	0.0143	0.0143	0.0143	0.0429	1.0000	
P. notoginseng	0.0143	0.0143	0.0143	0.0143	0.0143	0.0143	0.0429	0.2143	1.0000

[1)]*Panax ginseng* cultivars or accessions.

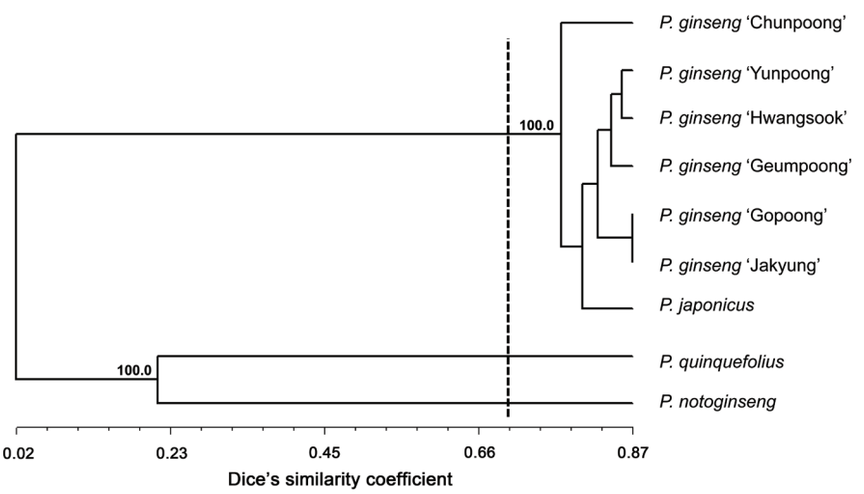

Fig. 3. Dendrogram of the nine *Panax* accessions, six *P. ginseng* accessions and three relative species. Phylogenetic tree was constructed based on the genotypes of 70 markers using unweighted pair group method with the arithmetic mean clustering analysis. Bootstrap values were calculated by 1,000 replications and only significant values were denoted on the branches.

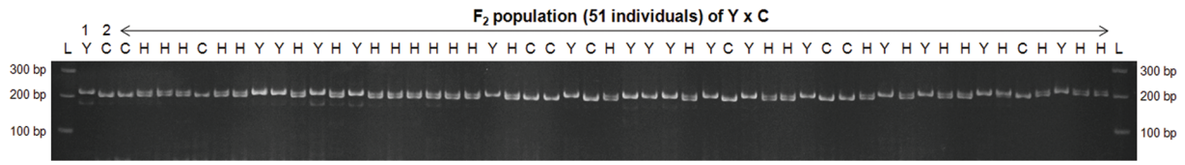

Fig. 4. Segregation of the polymorphic marker GES0019 in a F_2 population between Yunpoong and Chunpoong. polymerase chain reaction products were separated by non-denaturing polyacrylamide gel electrophoresis. Lanes and L indicate DNA ladder; 1, Yunpoong; 2, Chunpoong. F_2 population includes 51 F_2 individuals. Y, C, and H indicate genotype of each F_2 individual which is same with Yunpoong, Chunpoong, and Yunpoong/Chunpoong heterozygote, respectively. GES, ginseng expressed sequence tag-simple sequence repeat.

notable that *P. japonicus* clustered among the *P. ginseng* accessions with high similarity coefficients which averaged 0.7905, and 94.3% of the alleles of *P. japonicus* were observed in six *P. ginseng* accessions. Similarity coefficients among *P. ginseng* accessions ranged from 0.7571 to 0.8714 (Table 3). *P. notoginseng* and *P. quinquefolius* were divided with *P. ginseng*-*P. japonicus* clade with similarity coefficients of 0.0209 and 0.0207, respectively. The similarity coefficient between *P. notoginseng* and *P. quinquefolius* was 0.2149 which is higher than their value with *P. ginseng*-*P. japonicus* clade. The coefficient value is not coincided with the transferability value of each marker. Even though 97.1% of markers were amplified in *P. quinquefolius*, only 75.7% were amplified in *P. notoginseng* that indicate *P. ginseng*-*P. japonicus* clade is much closer to *P. quinquefolius* than to *P. notoginseng*. The biased data might be derived from genotype scoring method because only band appearance was counted and non-amplification was treated as missing. Sequence level analyses will clearly show the phylogenetic relationships of the species such as several studies based on conserved DNA sequences such as internal transcribed spacer sequences [59-62] and chloroplast DNA [60,63,64].

Reproducibility and utility of the markers

Most of the markers reported in the ginseng genome were limited to identification of individuals instead of authentication of cultivars or accessions because of the difficulty of genetic studies and the limited utility of pure inbred lines. Therefore, no inheritance study has yet been reported in ginseng. Our purpose was to develop stable and reproducible polymorphic markers which can discriminate elite cultivars and can be used for genetic mapping. Therefore, we have selected polymorphic markers between DNA pools of 15 individual plants for representing each accession to identify major polymorphic markers by overcoming the heterogeneity. Furthermore, to determine stable and reproducible inheritance of the markers, we analyzed seven of the GES markers against 51 F_2 individuals resulting from a cross between 'Yunpoong' and 'Chunpoong' which were most diverse elite ginseng cultivars (Fig. 4). All the markers segregated with a good fit to the Mendelian 1:2:1 ratio for the genotype of Yunpoong homozygote:heterozygote:Chunpoong homozygote (Table 4 and Fig. 4) indicating that these inheritable and reproducible markers can be utilized for discrimination of each cultivar and for mapping using the F_2 population.

Table 4. Goodness-of-fit analysis for seven markers in a F_2 population between a cross of 'Yunpoong' x 'Chunpoong'

Marker	Observed value			χ^2-value	*p*-value
	Yunpoong	Heterozygote	Chunpoong		
GES0003	10	30	11	1.63	0.44
GES0010	11	28	12	0.53	0.77
GES0013	18	24	9	3.35	0.19
GES0014	14	24	13	0.22	0.90
GES0015	18	26	7	4.76	0.09
GES0018	12	23	16	1.12	0.57
GES0019	16	26	9	1.94	0.38

GES, ginseng expressed sequence tag-simple sequence repeat.

There is no report on availability of segregating population in *P. ginseng* because of lack of pure inbred lines and genetic study. Both parental lines, 'Yunpoong' and 'Chunpoong', showed relatively high homogeneous genotypes with less than 10% of off-type alleles. Furthermore, both cultivars show distinct agricultural characteristics such as stem numbers, root shapes, disease durability and fruit colors. 'Yunpoong' is known as a best cultivar for high yield of roots with vigorous growth and 'Chunpoong' is known as a best cultivar for red ginseng processing [4]. We have identified 14 polymorphic markers between these two cultivars and seven of them showed clear mappable genotype scores for each F_2 individual. Recently, we had obtained a large scale transcriptome sequence data from both parental lines using Roche GS FLX Titanium platform and 50 Gbp of whole genome sequencing data from one of the parental line, 'Chunpoong', using Illumina Genome Analyzer II plat-

form [65]. Application of fast evolving next generation sequencing technology and the utility of the mapping population may promise acceleration of high density genetic mapping and complete genome sequencing of the mysterious medicinal plant, ginseng.

ACKNOWLEDGEMENTS

This study was supported by the Mid-career Researcher Program through a National Research Foundation of Korea grant funded by the Ministry of Education, Science and Technology (no. R01-2007-000-20823-0) and a grant from the Next-Generation BioGreen 21 Program (no. PJ008202), Rural Development Administration, Republic of Korea.

REFERENCES

1. Yun TK. Brief introduction of *Panax ginseng* C.A. Meyer. J Korean Med Sci 2001;16 Suppl:S3-S5.
2. Attele AS, Wu JA, Yuan CS. Ginseng pharmacology: multiple constituents and multiple actions. Biochem Pharmacol 1999;58:1685-1693.
3. Choi KT, Kim YT, Kwon WS. Present status in development of new ginseng varieties. Korean J Ginseng Sci 1992;16:164-168.
4. Lee JS, Lee SS, Lee JS, Ahn IO. Effect of seed size and cultivars on the ratio of seed coat dehiscence and seedling performance in *Panax ginseng*. J Ginseng Res 2008;32:257-263.
5. Artiukova EV, Kozyrenko MM, Reunova GD, Muzarok TI, Zhuravlev IuN. Analysis of genomic variability of planted *Panax ginseng* by RAPD. Mol Biol (Mosk) 2000;34:339-344.
6. Kim CK, Choi HK. Genetic diversity and relationship in Korean ginseng (*Panax schinseng*) based on RAPD analysis. Korean J Genet 2003;25:181-188.
7. Ma XJ, Wang XQ, Xu ZX, Xiao PG, Hong DY. RAPD variation within and among populations of ginseng cultivars. Acta Bot Sin 2000;42:587-590.
8. Ha WY, Shaw PC, Liu J, Yau FC, Wang J. Authentication of *Panax ginseng* and *Panax quinquefolius* using amplified fragment length polymorphism (AFLP) and directed amplification of minisatellite region DNA (DAMD). J Agric Food Chem 2002;50:1871-1875.
9. Kim J, Jo BH, Lee KL, Yoon ES, Ryu GH, Chung KW. Identification of new microsatellite markers in *Panax ginseng*. Mol Cells 2007;24:60-68.
10. Ma KH, Dixit A, Kim YC, Lee DY, Kim TS, Cho EG, Park YJ. Development and characterization of new microsatellite markers for ginseng (*Panax ginseng* C. A. Meyer). Conserv Genet 2007;8:1507-1509.
11. Van Dan N, Ramchiary N, Choi SR, Uhm TS, Yang TJ, Ahn IO, Lim YP. Development and characterization of new microsatellite markers in *Panax ginseng* (C.A. Meyer) from BAC end sequences. Conserv Genet 2010;11:1223-1225.
12. Choi DW, Jung J, Ha YI, Park HW, In DS, Chung HJ, Liu JR. Analysis of transcripts in methyl jasmonate-treated ginseng hairy roots to identify genes involved in the biosynthesis of ginsenosides and other secondary metabolites. Plant Cell Rep 2005;23:557-566.
13. Jung JD, Park HW, Hahn Y, Hur CG, In DS, Chung HJ, Liu JR, Choi DW. Discovery of genes for ginsenoside biosynthesis by analysis of ginseng expressed sequence tags. Plant Cell Rep 2003;22:224-230.
14. Kim MK, Lee BS, In JG, Sun H, Yoon JH, Yang DC. Comparative analysis of expressed sequence tags (ESTs) of ginseng leaf. Plant Cell Rep 2006;25:599-606.
15. Sathiyamoorthy S, In JG, Gayathri S, Kim YJ, Yang DC. Generation and gene ontology based analysis of expressed sequence tags (EST) from a *Panax ginseng* C. A. Meyer roots. Mol Biol Rep 2010;37:3465-3472.
16. Nagaraj SH, Gasser RB, Ranganathan S. A hitchhiker's guide to expressed sequence tag (EST) analysis. Brief Bioinform 2007;8:6-21.
17. Varshney RK, Graner A, Sorrells ME. Genic microsatellite markers in plants: features and applications. Trends Biotechnol 2005;23:48-55.
18. Allen GC, Flores-Vergara MA, Krasynanski S, Kumar S, Thompson WF. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nat Protoc 2006;1:2320-2325.
19. Kofler R, Torres TT, Lelley T, Schlotterer C. PanGEA: identification of allele specific gene expression using the 454 technology. BMC Bioinformatics 2009;10:143.
20. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 1999;27:573-580.
21. Conesa A, Gotz S, García-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 2005;21:3674-3576.
22. Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 2005;21:2128-2129.
23. Dice LR. Measures of the amount of ecologic association between species. Ecology 1945;26:297-302.
24. Sneath PH, Sokal RR. Numerical taxonomy: the principles and practice of numerical classification. San Francisco: W. H. Freeman, 1973.

25. Yap IV, Nelson RJ. Winboot: a program for performing bootstrap analysis of binary data to determine the confidence limits of UPGMA-based dendrograms. Manila: IRRI, 1996.
26. Clarke LA, Rebelo CS, Goncalves J, Boavida MG, Jordan P. PCR amplification introduces errors into mononucleotide and dinucleotide repeat sequences. Mol Pathol 2001;54:351-353.
27. Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L. Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. Theor Appl Genet 2007;114:359-372.
28. Fraser LG, Harvey CF, Crowhurst RN, De Silva HN. EST-derived microsatellites from Actinidia species and their potential for mapping. Theor Appl Genet 2004;108:1010-1016.
29. Rungis D, Berube Y, Zhang J, Ralph S, Ritland CE, Ellis BE, Douglas C, Bohlmann J, Ritland K. Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. Theor Appl Genet 2004;109:1283-1294.
30. Cloutier S, Niu Z, Datla R, Duguid S. Development and analysis of EST-SSRs for flax (*Linum usitatissimum* L.). Theor Appl Genet 2009;119:53-63.
31. Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ. Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. Plant Sci 2001;160:1115-1123.
32. Kantety RV, La Rota M, Matthews DE, Sorrells ME. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 2002;48:501-510.
33. Nicot N, Chiquet V, Gandon B, Amilhat L, Legeai F, Leroy P, Bernard M, Sourdille P. Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). Theor Appl Genet 2004;109:800-805.
34. Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ. Analysis of SSRs derived from grape ESTs. Theor Appl Genet 2000;100:723-726.
35. Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 2003;106:411-422.
36. Varshney RK, Thiel T, Stein N, Langridge P, Graner A. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 2002;7:537-546.
37. Zhuravlev YN, Reunova GD, Kats IL, Muzarok TI, Bondar AA. Genetic variability and population structure of endangered *Panax ginseng* in the Russian Primorye. Chin Med 2010;5:21.
38. Cruse-Sanders JM, Hamrick JL. Genetic diversity in harvested and protected populations of wild American ginseng, *Panax quinquefolius* L. (Araliaceae). Am J Bot 2004;91:540-548.
39. Mooney EH, Mcgraw JB. Effects of self-pollination and outcrossing with cultivated plants in small natural populations of American ginseng, *Panax quinquefolius* (Araliaceae). Am J Bot 2007;94:1677-1687.
40. Wang D, Hong D, Koh HL, Zhang YJ, Yang CR, Hong Y. Biodiversity in cultivated *Panax notoginseng* populations. Acta Pharmacol Sin 2008;29:1137-1140.
41. Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Genet Genomics 2003;270:315-323.
42. Liang X, Chen X, Hong Y, Liu H, Zhou G, Li S, Guo B. Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and Arachis wild species. BMC Plant Biol 2009;9:35.
43. Saha MC, Mian MA, Eujayl I, Zwonitzer JC, Wang L, May GD. Tall fescue EST-SSR markers with transferability across several grass species. Theor Appl Genet 2004;109:783-791.
44. Yu JK, Dake TM, Singh S, Benscher D, Li WL, Gill B, Sorrells ME. Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. Genome 2004;47:805-818.
45. Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W. Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 2002;104:399-407.
46. Han ZG, Guo WZ, Song XL, Zhang TZ. Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton. Mol Genet Genomics 2004;272:308-327.
47. Hisano H, Sato S, Isobe S, Sasamoto S, Wada T, Matsuno A, Fujishiro T, Yamada M, Nakayama S, Nakamura Y *et al*. Characterization of the soybean genome using EST-derived microsatellite markers. DNA Res 2007;14:271-281.
48. Varshney RK, Grosse I, Hahnel U, Siefken R, Prasad M, Stein N, Langridge P, Altschmied L, Graner A. Genetic mapping and BAC assignment of EST-derived SSR markers shows non-uniform distribution of genes in the barley genome. Theor Appl Genet 2006;113:239-250.
49. Chabane K, Ablett GA, Cordeiro GM, Valkoun J, Henry RJ. EST versus genomic derived microsatellite markers for genotyping wild and cultivated barley. Genet Resour

Crop Evol 2005;52:903-909.
50. Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 2000;100:713-722.
51. Eujayl I, Sorrells M, Baum M, Wolters P, Powell W. Assessment of genotypic variation among cultivated durum wheat based on EST-SSRS and genomic SSRS. Euphytica 2001;119:39-43.
52. Russell J, Booth A, Fuller J, Harrower B, Hedley P, Machray G, Powell W. A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley genome. Genome 2004;47:389-398.
53. Hong CP, Lee SJ, Park JY, Plaha P, Park YS, Lee YK, Choi JE, Kim KY, Lee JH, Lee J *et al*. Construction of a BAC library of Korean ginseng and initial analysis of BAC-end sequences. Mol Genet Genomics 2004;271:709-716.
54. Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C *et al*. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 2007;449:463-467.
55. Yang TJ, Kim JS, Kwon SJ, Lim KB, Choi BS, Kim JA, Jin M, Park JY, Lim MH, Kim HI *et al*. Sequence-level analysis of the diploidization process in the triplicated *FLOWERING LOCUS C* region of *Brassica rapa*. Plant Cell 2006;18:1339-1347.
56. Bulgakov VP, Lauve LS, Chernoded GK, Khodakovskaia MV, Zhuravlev IuN. Chromosomal variability of ginseng cells transformed with plant oncogene rolC. Genetika 2000;36:209-216.
57. Besnard G, Garcia-Verdugo C, De Casas RR, Treier UA, Galland N, Vargas P. Polyploidy in the olive complex (*Olea europaea*): evidence from flow cytometry and nuclear microsatellite analyses. Ann Bot 2008;101:25-30.
58. Grushvitskii IV. Zhen'shen : Voprosy biologii. Leningrad: Akad. Nauk SSSR, 1961.
59. Artiukova EV, Goncharov AA, Kozyrenko MM, Reunova GD, Zhuravlev IuN. Phylogenetic relationships of the Far Eastern Araliaceae inferred from ITS sequences of nuclear rDNA. Genetika 2005;41:800-810.
60. Choi HK, Wen J. A phylogenetic analysis of *Panax* (Araliaceae): integrating cpDNA restriction site and nuclear rDNA ITS sequence data. Plant Syst Evol 2000;224:109-120.
61. Wen J, Plunkett GM, Mitchell AD, Wagstaff SJ. The evolution of Araliaceae: a phylogenetic analysis based on ITS sequences of nuclear ribosomal DNA. Syst Bot 2001;26:144-167.
62. Wen J, Zimmer EA. Phylogeny and biogeography of *Panax* L (the ginseng genus, Araliaceae): inferences from ITS sequences of nuclear ribosomal DNA. Mol Phylogenet Evol 1996;6:167-177.
63. Lee C, Wen J. Phylogeny of *Panax* using chloroplast *trn*C-*trn*D intergenic region and the utility of *trn*C-*trn*D in interspecific studies of plants. Mol Phylogenet Evol 2004;31:894-903.
64. Zhu S, Fushimi H, Cai S, Komatsu K. Phylogenetic relationship in the genus *Panax*: inferred from chloroplast *trn*K gene and nuclear 18S rRNA gene sequences. Planta Med 2003;69:647-653.
65. Choi HI, Kim NH, Jung JY, Park HM, Park HS, Park JY, Lee J, Park J, Lee J, Choi BS *et al*. Current status of Korean ginseng (*Panax ginseng*) genome mapping and sequencing. In: Yang DC, Kim SK, Oh HI, eds. Advances in ginseng research. Proceedings of 10th International Symposium on Ginseng; 2010 Sep 13-16; Seoul, Korea. Seoul: The Korean Society of Ginseng, 2010. p. 762-778.