Indonesian Journal of Agricultural Science Vol. 17 No. 1 April: 17–26 DOI: http//dx.doi.org/10.21082/ijas.v17n1.2016.p.17-26

DEVELOPMENT OF EST-SSR MARKERS TO ASSESS GENETIC DIVERSITY OF BROCCOLI AND ITS RELATED SPECIES

Pengembangan Marka EST-SSR untuk Analisis Keragaman Genetik Tanaman Brokoli dan Kerabatnya

Nur Kholilatul Izzah[a]*, Reflinur[b] and Tae-Jin Yang[c]

[a]*Indonesian Industrial and Beverage Crops Research Institute*
Jalan Raya Pakuwon km 2, Sukabumi 43357, West Java
Phone +62 266 6542181, Fax. +62 266 6542087, E-mail: balittri@litbang.pertanian.go.id
[b]*Indonesian Center of Agricultural Biotechnology and Genetic Resources Research and Development*
Jalan Tentara Pelajar No. 3A, Bogor 16111, West Java
[c]*Department of Plant Science and Research Institute for Agriculture and Life Sciences, Seoul National University*
Seoul, 151-921, Korea
**Corresponding author: lila_ref@yahoo.co.id*

Submitted 15 August 2015; Revised 10 March 2016; Accepted 15 March 2016

ABSTRACT

Development of Expressed Sequence Tag-Simple Sequence Repeat (EST-SSR) markers derived from public database is known to be more efficient, faster and low cost. The objective of this study was to generate a new set of EST-SSR markers for broccoli and its related species and their usefulness for assessing their genetic diversity. A total of 202 *Brassica oleracea* ESTs were retrieved from NCBI and then assembled into 172 unigenes by means of CAP3 program. Identification of SSRs was carried out using web-based tool, RepeatMasker software. Afterwards, EST-SSR markers were developed using Primer3 program. Among the identified SSRs, trinucleotide repeats were the most common repeat types, which accounted for about 50%. A total of eight primer pairs were successfully designed and yielded amplification products. Among them, five markers were polymorphic and displayed a total of 30 alleles with an average number of six alleles per locus. The polymorphic markers were subsequently used for analyzing genetic diversity of 36 *B. oleracea* cultivars including 22 broccoli, five cauliflower and nine kohlrabi cultivars based on genetic similarity matrix as implemented in NTSYS program. At similarity coefficient of 61%, a UPGMA clustering dendrogram effectively separated 36 genotypes into three main groups, where 30 out of 36 genotypes were clearly discriminated. The result obtained in the present study would help breeders in selecting parental lines for crossing. Moreover, the novel EST-SSR markers developed in the study could be a valuable tool for differentiating cultivars of broccoli and related species.

[***Keywords*:** Broccoli, cauliflower, kohlrabi, EST-SSR markers, genetic diversity]

ABSTRAK

Pengembangan marka Expressed Sequence Tag-Simple Sequence Repeat *(EST-SSR) yang berasal dari database publik dikenal lebih efisien, cepat, dan berbiaya rendah. Tujuan penelitian ini adalah untuk mendapatkan satu set marka EST-SSR baru untuk brokoli dan kerabatnya, serta kegunaannya untuk analisis keragaman genetik. Sebanyak 202 sekuen EST dari* Brassica oleracea *yang diperoleh dari NCBI digunakan dalam penelitian ini, yang selanjutnya dikelompokkan menjadi 172 unigenes dengan menggunakan program CAP3. Identifikasi motif SSR dilakukan menggunakan program RepeatMasker, sedangkan marka EST-SSR didesain menggunakan program Primer3. Sebanyak 12 SSR berhasil dideteksi dan di antara SSR yang teridentifikasi, trinukleotida merupakan jenis pengulangan yang paling banyak ditemukan, yaitu sekitar 50%. Selain itu, delapan pasang primer berhasil didesain dan menghasilkan produk amplifikasi. Hasil amplifikasi menunjukkan lima primer bersifat polimorfis dan menghasilkan 30 alel dengan rata-rata enam alel per lokus. Marka polimorfis tersebut kemudian digunakan untuk menganalisis keragaman genetik 36 kultivar* B. oleracea*, yang meliputi 22 brokoli, 5 kol bunga, dan 9 kohlrabi berdasarkan matriks kemiripan genetik seperti yang diterapkan dalam program NTSYS. Pada koefisien kesamaan 61%, UPGMA dendrogram berhasil mengelompokkan 36 genotipe menjadi tiga kelompok utama. Sebanyak 30 dari 36 genotipe tersebut berhasil dibedakan satu dengan yang lainnya. Hasil yang diperoleh dalam penelitian ini dapat membantu para pemulia dalam memilih tetua yang digunakan untuk persilangan. Selain itu, marka EST-SSR yang didesain pada penelitian ini merupakan suatu alat yang berharga untuk membedakan kultivar-kultivar brokoli dan kerabatnya.*

[***Kata kunci*:** Brokoli, kol bunga, kohlrabi, Marka EST-SSR, keragaman genetik]

INTRODUCTION

Broccoli (*Brassica oleracea* L. var. *italica* Plenck) is one of the most important vegetables in the world. Broccoli has noteworthy nutrients including vitamin C, vitamin A (mostly as beta-carotene), folic acid, calcium and fiber. The popularity of broccoli has

been steadily increasing in many countries that led to the rising number of new cultivars. Characterization of many different cultivars produced by different world-wide companies would provide valuable information for the introduction and genetic improvement of broccoli cultivars (Lu *et al.* 2009).

One of the main interests in broccoli breeding program is to develop new varieties resistant to black rot disease caused by *Xanthomonas campestris* pv. *campestris* (Pammel) Dowson (*Xcc*). Black rot disease is widely known as the most destructive disease in broccoli and its related species, which has a broad range geographical distribution (Meenu *et al*. 2013). Development of resistant varieties can be achieved by characterization of broccoli germplasm collection through morphological and molecular approaches.

Many scientists exert a lot of efforts to obtain new source of black rot resistance through application of molecular markers. Sharma *et al*. (2016) utilized 364 intron length polymorphic and microsatellite markers to find new gene source for black rot resistance. Meanwhile, Lee *et al*. (2015) successfully identified four QTLs related to black rot resistance. Hence, development of molecular markers related to black rot disease is important to enhance breeding program in broccoli and its related species.

Molecular markers have been proven to be a powerful tool for many genetic analyses. One of the advantage of molecular markers is more reliable than morphological characteristics since they are not affected by environmental factors (Ali *et al.* 2007). During the last two decades, the invention of polymerase chain reaction (PCR) has generated a variety of PCR-based markers including random amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLP), sequence-related amplified polymorphisms (SRAP), simple sequence repeats (SSRs) and single nucleotide polymorphism (SNP). Among those markers, SSRs become marker of choice because it requires only a small amount of DNA, easily detectable by PCR, co-dominantly inherited, multi-allelic, abundant and amenable to high-throughput analysis (Kalia *et al*. 2011).

SSRs, also known as microsatellites, are tandem repeats of 2–6 bp DNA core sequences that are widely distributed in both non-coding and transcribed sequences, generally known as genomic-SSRs and EST-SSRs, respectively (Shirasawa *et al.* 2011). Genomic SSRs are highly polymorphic and widely distributed throughout the genome (Wang *et al.* 2011). However, development of genomic SSR markers is costly, laborious and time-consuming because it requires a small-insert genomic library and performs hybridization with SSR oligonucleotides and sequence candidate clones (Yi *et al*. 2006). Fortunately, the availability of EST sequences in public databases overcomes the difficulty in developing genomic SSR markers through conventional method. ESTs are particularly attractive for marker development since they represent coding regions of the genome, and are also being developed at an extremely fast pace for many genomes (Kumpatla and Mukhopadhyay 2005). Therefore, exploitation of EST-SSRs has become more feasible, low cost and faster by means of bio-informatics tools such as Sputnik, SSRFinder, MIcroSAtellite (MISA), Tandem Repeat Finder (TRF) (Varshney *et al*. 2005) and SSRIT (Temnykh *et al.* 2001). With these tools, SSRs are easily obtained by electronic search of EST databases (Hu *et al.* 2010).

Up to now, EST-SSRs have been successfully identified and developed in various plant species, and used for multiple applications including genetic diversity analysis, cultivar identification, construction of high-density genetic linkage maps, marker-assisted selection, identification of quantitative trait loci (QTLs), association mapping and molecular breeding (Kumari *et al.* 2013). EST-SSRs, which located in coding region of the genome, demonstrated some valuable advantages such as they can be rapidly found by electronic sorting and have greater transferability between species than genomic SSR, since genic regions are more conserved among related species (Shirasawa *et al*. 2011; Chen *et al*. 2015). Furthermore, EST-SSRs usually present in gene-rich regions and can be used as anchor markers for comparative mapping and genetic evolutionary studies (Zhou *et al*. 2014).

The great advantages of EST-SSR markers have attracted the attention of many scientists, which then use these markers for genetic diversity study. For instance, Wen *et al*. (2010) had been successfully used 36 EST-SSR markers for classifying 45 *Jatropha curcas* accessions into six groups, which showed correlation with geographic origin. A recent study reported a set of EST-SSR and gSSR markers which clearly separated 91 commercial *B. oleracea* cultivars belonging to six subspecies including cabbage, broccoli, cauliflower, kohlrabi, kale and kailan into six different clusters with a tendency to cluster into its subspecies (Izzah *et al.* 2013). Moreover, another research done by Zhao *et al*. (2014) also exhibited the power of SSR markers in dividing 49 cauliflower genotypes into three major groups based on their curd types. In present study, we report on the

development of novel EST-SSR markers using *B. oleracea* SSH-cDNA library of black rot infection derived from public database, and their utility for assessing genetic diversity of broccoli and its related species. In addition, these newly developed markers will also be useful for further breeding program in broccoli and its related species, especially to develop black rot resistant varieties.

MATERIALS AND METHODS

Plant Materials and DNA Extraction

A total of 36 genotypes consisting of 22 broccoli cultivars (*B. oleracea* var. *italica* Plenck), five cauliflowers (*B. oleracea* var. *botrytis*), and nine kohlrabies (*B. oleracea* var. *gongylodes*) (Table 1)

Table 1. List thirty six of broccoli, cauliflower and kohlrabi cultivars used in this study and their phenotypic characteristics.

Cultivar Code	Cultivar name	Varietal group	Source	Characteristics of main phenotype
B2008	Yuan you qing hua cai	Broccoli	Tokita (Japan)	Early maturity, domed head shape, medium bead size
B2013	Yu huang	Broccoli	Hongkong Seed (Japan)	Medium maturity, domed head shape, medium bead size
B2014	Youshou	Broccoli	Sakata (Japan)	Early maturity, domed head shape, fine bead size
B2056	Heart Land	Broccoli	Sakata (Japan)	Medium maturity, domed head shape, anthocyanin-free, medium bead size
B2060	Subaru	Broccoli	Brolead (Japan)	Early maturity, domed head shape, fine bead size
B2061	Fighter	Broccoli	Brolead (Japan)	Early maturity, domed head shape, anthocyanin-free, fine bead size
B2065	KB-052	Broccoli	Mikado-Kyowa (Japan)	Early maturity, fine bead size
B2070	Green Magic	Broccoli	Sakata (Japan)	Early maturity, domed head shape, fine bead size
B2071	Tradition	Broccoli	Seminis (U.S.A.)	Early maturity, domed head shape, fine bead size
B2073	Montop	Broccoli	Syngenta (Switzerland)	Early maturity, domed head shape, fine bead size
B2085	Green Belt	Broccoli	Sakata (Japan)	Medium maturity, domed head shape, medium bead size
B2097	Grace	Broccoli	Bejo (Netherlands)	Medium maturity, domed head shape, medium bead size
B2098	Super Grace	Broccoli	Bejo (Netherlands)	Medium maturity, domed head shape, medium bead size
B2134	Castle	Broccoli	Takii (Japan)	Early maturity, fine bead size
B2135	Anfree-747	Broccoli	Takii (Japan)	Early maturity, fine bead size, anthocyanin-free
B2138	Marathon	Broccoli	Sakata (Japan)	Late maturity, high domed head shape, fine bead size
B2139	BI-15 (Monaco)	Broccoli	Syngenta (Switzerland)	Late maturity, high domed head shape, fine bead size
B2140	Heritage	Broccoli	Seminis (U.S.A.)	Late maturity, high domed head shape, fine bead size
B2145	Ironman	Broccoli	Seminis (Netherlands)	Late maturity, high domed head shape, fine bead size
B2193	Aosima	Broccoli	Sakata (Japan)	Late maturity, high domed head shape, anthocyanin-free, fine bead size
B2198	Green Dome	Broccoli	Takii (Japan)	Late maturity, domed head shape, anthocyanin-free, fine bead size
B2205	Endevour	Broccoli	Takii (Japan)	Late maturity, domed head shape, anthocyanin-free, fine bead size
B2266	Snow Dream	Cauliflower	Takii (Japan)	Medium maturity, white curd color, high domed head shape, good coverage
B2267	White Dream	Cauliflower	Takii (Japan)	Medium maturity, white curd color, high domed head shape, good coverage
B2268	Snow March	Cauliflower	Takii (Japan)	Medium maturity, white curd color, high domed head shape, good coverage
B2270	Violet Dream	Cauliflower	Takii (Japan)	Early maturity, violet curd color, early bolting type
B2271	Orange Dream	Cauliflower	Takii (Japan)	Medium maturity, orange curd color, high domed head shape
K3001	Korist	Kohlrabi	Bejo (Netherlands)	Early maturity, round head shape, milky skin color
K3008	Express Forcer	Kohlrabi	Takii (Japan)	Early maturity, flat head shape, pale green color
K3038	White Rookie	Kohlrabi	Numhems Korea (Korea)	Early maturity, flat head shape, green color
K3039	Winner	Kohlrabi	Takii (Japan)	Medium maturity, flat head shape, pale green color
K3044	UFO	Kohlrabi	Seminis (Korea)	Early maturity, flat head shape, green color
K3048	Worldcol	Kohlrabi	Joeun Seed (Korea)	Early maturity, flat head shape, green color, very long field holding ability
K3065	Kolibri	Kohlrabi	Bejo (Netherlands)	Early maturity, flat head shape, red color
K3066	Purple King	Kohlrabi	Joeun Seed (Korea)	Early maturity, flat head shape, red color, less fiber
K3083	Dongchuan	Kohlrabi	Konmyeong Noksaeng Chaejo Seed (China)	Late maturity, flat head shape, green color, high fiber, early bolting type

were used for polymorphic screening of newly designed EST-SSR markers, and subsequently used for genetic diversity analysis. All of plant materials examined in the present study were kindly provided by Joeun Seed Company, Korea. Genomic DNA was extracted from fresh young leaf tissue according to the modified cetyltrimethylammonium bromide (CTAB) method (Allen *et al.* 2006). Measurement of the quality and quantity of the extracted DNA were done using a NanoDrop ND-1000 (NanoDrop Technologies, Inc., Wilmington, DE, USA). The final concentration of each DNA sample was adjusted to 10 ng μl^{-1} for PCR analysis.

EST-SSR Markers Development

A total of 202 EST sequences of *B. oleracea* SSH-cDNA library of black rot infection in leaves of resistant variety compared with its susceptible near-isogenic line were downloaded from NCBI database (http://www.ncbi.nlm.nih.gov/). These ESTs were then assembled using CAP3 program with the following criteria: 93% identity and 40 bp overlap (Huang and Madan, 1999). A web-based tool, RepeatMasker software (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker) was used for searching SSRs in the unigenes with the criteria as follows: a minimum of seven repeats for dinucleotide motifs, six repeats for trinucleotide motifs, and five repeats for tetra-, penta- and hexanucleotide motifs. These SSRs were then selected based on the length of flanking sequence. SSRs with long flanking sequence were subsequently used for designing primer; in contrast, SSRs which have short flanking sequence were removed due to insufficient for primer design. Primer3 program (http://bioinfo.ut.ee/primer3–0.4.0/) was used for designing primers from the flanking sequence of each SSR motif. The input criteria for primer design were as follows: a length of 18–24 bp, a GC content of 40–80%, an estimated amplicon size of 150–350 bp and melting temperatures of 55–63° C. EST-SSR markers developed in the present study were designated as "BoEMS", representing ***B****rassica* ***o****leracea* **E**ST **m**icro**s**atellites. All of the primers were synthesized by Macrogen (Seoul, Korea).

Amplification of EST-SSR Markers

PCR amplifications were performed in a total volume of 10 µl containing 10 ng DNA template, 1 x PCR reaction buffer (Inclone Biotech), 0.2 mM each dNTP (Inclone Biotech), 0.2 µM each primer and 1 unit *Taq* DNA polymerase (Inclone Biotech). The reaction mixture was initially denatured at 94° C for 4 minutes, followed by 35 cycles of amplifications at 94° C for 30 seconds, 54° C for 30 seconds and 72° C for 30 seconds, and a final extension at 72° C for 10 minutes. The PCR products were separated on 6% non-denaturing polyacrylamide gel electrophoresis using 1 x TBE buffer. The gels were stained with ethidium bromide for 20 minutes and DNA bands were visualized under UV light using the gel documentation system.

Data Analysis

The reproducible amplified fragments were scored "1" to indicate the presence and "0" to indicate the absence of a band. The genetic relationship among 36 cultivars was evaluated based on genetic similarity coefficient as implemented in NTSYS-PC version 2.1 (Rohlf 2000). Clustering analysis was carried out using the unweighted pair group arithmetic mean method (UPGMA) in the SAHN subprogram of NTSYS-PC. The similarity matrix was also used for principal coordinate analysis (PCoA).

PowerMarker version 3.25 (Liu and Muse 2005) was also used for calculating number of alleles (N_A), specific alleles (S_A), major allele frequency (M_{AF}), gene diversity (GD), expected heterozygosity (H_e) and polymorphic information content (PIC) values. Major allele frequency (M_{AF}) was defined as allele with the highest frequency.

RESULTS

Characteristics of SSRs from ESTs in *B. oleracea*

In order to facilitate the development of novel EST-SSR markers for broccoli, we analyzed 202 EST sequences from *B. oleracea* SSH-cDNA library of black rot infection in leaves of resistant variety compared with its susceptible near-isogenic line, representing a total length of 68,925 kb. These EST sequences were assembled into 172 unigenes consisting of 20 contigs and 152 singletons (Fig. 1). Of those unigenes, we identified 12 EST sequences containing SSRs (SSR-ESTs). Analysis of the nucleotide sequences of EST containing SSRs exhibited that trinucleotide was the most abundant type, represented by 50% among identified SSRs,

whereas other SSR types including dinucleotide, hexanucleotide and heptanucleotide repeats accounted for 16.67% each (Table 2). Regarding to repeat motifs, CTT motifs which member of trinucleotide repeat exhibited higher frequency (33.33%) compared to the other repeat motifs.

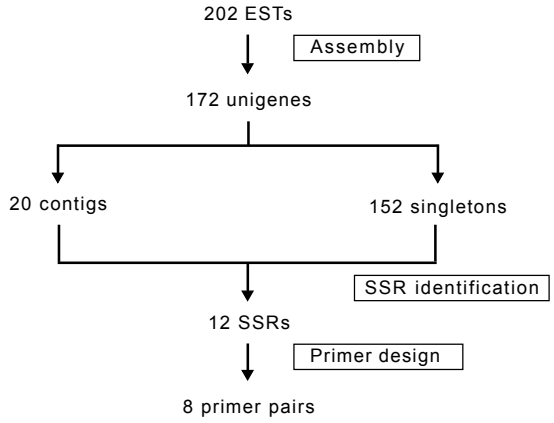

Fig. 1. Diagram illustrating the EST assembly and SSR identification.

Table 2. Characterization of SSRs identified in 202 *Brassica oleracea* SSH-cDNA library EST.

Repeat motif	Repeat number								Total	%
	3	4	5	6	7	8	9	10		
Dinucleotide	-	-	-	-	-	1	1	-	2	16.67
Trinucleotide	-	-	1	2	1	1	-	1	6	50
Hexanucleotide	2	-	-	-	-	-	-	-	2	16.67
Heptanucleotide	2	-	-	-	-	-	-	-	2	16.67
Total	4	-	1	2	1	2	1	1	12	100

Development and Polymorphism Level of EST-SSR Markers

A total of 12 EST containing SSR (SSR-ESTs) were used for designing primer, which allowed the development of eight EST-SSR primer pairs. The remaining four SSR-ESTs could not be designed as primer due to short flanking sequence. These newly designed primers were then evaluated their applicability and polymorphism level by using 36 *B. oleracea* genotypes including 22 broccoli, five cauliflower and nine kohlrabi cultivars. All of the designed primers yielded unambiguous PCR products (Fig. 2). Furthermore, five of the eight EST-SSR markers revealed polymorphism among the 36 *B. oleracea* genotypes tested (Table 3).

Five polymorphic EST-SSR markers obtained in this study would be useful for identifying resistant cultivar since they related with black rot resistance. In addition, to know another function of these polymorphic markers, we also performed blast analysis against *Arabidopsis* protein sequences database using the BLASTX algorithm (http://www.ncbi.nlm.nih.gov/Blast). The best hits of the ESTs were assigned at expected value <10–6 (Table 4). The blast results showed that all of the polymorphic primers demonstrated significant hits to *Arabidopsis* protein sequences. This indicates that the polymorphic primers designed in the present study should be valuable for several genetic applications. Furthermore, these markers would also be helpful for further breeding program, particularly to develop resistant varieties against black rot disease in broccoli and its related species.

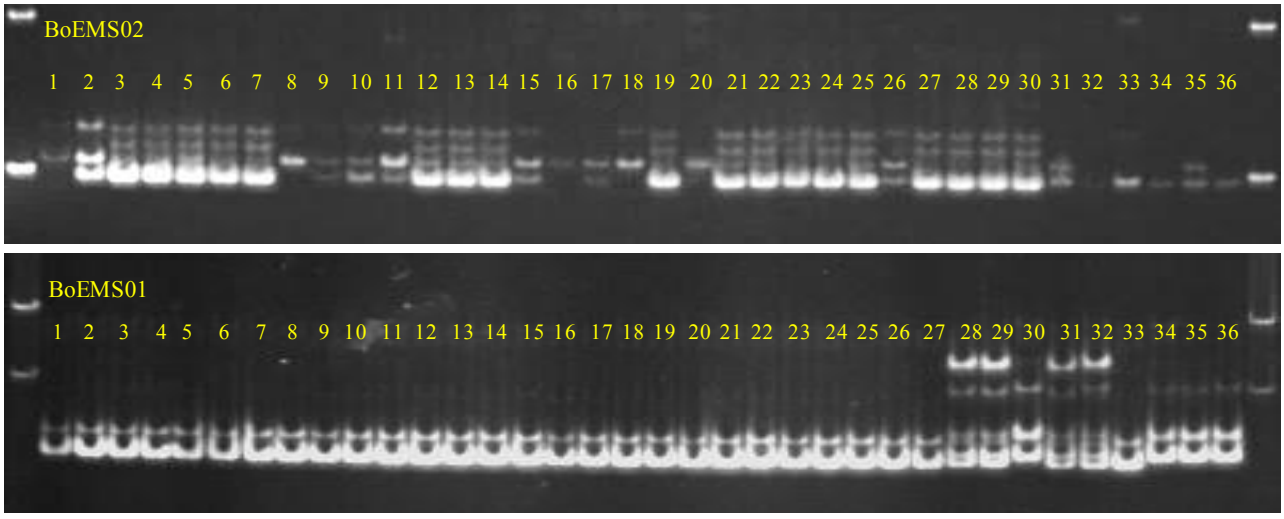

Fig. 2. Profile of two polymorphic EST-SSR markers separated using 6% non-denaturing polyacrylamide gel electrophoresis. 1–22: broccoli cultivars; 23–27: cauliflower cultivars; 28–36: kohlrabi cultivars.

Table 3. Features of polymorphic EST-SSR markers developed in present study.

Marker name	Motif	Forward sequence	Reverse sequence	Expected product size
BoEMS01	$(TA)_9$	CACCACACCCGATTTAAACA	AAAATCGGCTCAGACATTGC	191
BoEMS02	$(CTT)_6$	ATCGAAGCCAAGCTTTTGAA	ACACACCCAAGGCTTGTAGG	187
BoEMS05	$(TCA)_5$	CGCTTGGTTTAGCCTCTGAC	AGCTCCGTGCTCTTCTGGTA	268
BoEMS06	$(GGT)_7$	TTGACTTTGCAGCGCATAAC	GCTTTGCTTCATCAACACCA	261
BoEMS07	$(CTCCTT)_3$	GCGCATACTTCATCGACTCC	CTTGCAAATACGCCATCAGA	150

Table 4. Functional annotation of five polymorphic EST-SSR markers by BLASTX algorithm.

Marker name	Best hits	*Arabidopsis* Gene ID	E-value
BoEMS01	Light-harvesting chlorophyll protein complex II subunit B1	818006	2.00E-22
BoEMS02	Thioredoxin H4	838562	8.00E-09
BoEMS05	Chlorophyll a-b binding protein CP29.1	830325	3.00E-36
BoEMS06	ATP-dependent zinc metalloprotease FTSH 2	817646	9.00E-86
BoEMS07	Cell division control protein 48-A	820142	8.00E-45

Table 5. Summary statistics of five polymorphic EST-SSR markers across 36 *Brassica oleracea* cultivars.

Marker name	Allele number	Major allele frequency[1)] (%)	Gene diversity	H_o[2)]	PIC[3)]
BoEMS01	3	0.83	0.29	0.22	0.27
BoEMS02	6	0.48	0.69	0.80	0.65
BoEMS05	10	0.27	0.82	0.77	0.79
BoEMS06	8	0.47	0.72	0.78	0.69
BoEMS07	3	0.92	0.15	0.17	0.15
Mean	6	0.59	0.53	0.55	0.51

[1)]Major allele is defined as the allele showing the highest frequency.
[2)]Expected heterozygosity
[3)]Polymorphic information content

Assessment of Genetic Diversity in *B. oleracea* Cultivars

Five polymorphic EST-SSR markers were first analyzed using PowerMarker program to know their allelic diversity. The polymorphic markers produced a total of 30 alleles with an average of six alleles per locus (Table 5). The number of alleles for each locus ranged from three (BoEMS01 and BoEMS07) to ten (BoEMS05). Other components of allelic diversity such as frequency of major alleles, gene diversity, heterozygosity and PIC value on each locus were also investigated. The frequency of major alleles ranged from 0.27 (BoEMS05) to 0.92 (BoEMS07) with means of 0.59. An average of gene diversity across five loci was 0.53 with a range from 0.15 (BoEMS07) to 0.82 (BoEMS05). Meanwhile, an average of heterozygosity and PIC value was 0.55 and 0.51, respectively. Marker BoEMS02 exhibited the highest heterozygosity value (0.80), whereas the highest PIC value was shown by marker BoEMS05 (0.79). In contrast, marker BoEMS07 showed the lowest heterozygosity and PIC value of 0.17 and 0.15, respectively.

The genotyping data of the five polymorphic markers were then used for assessing the genetic diversity of broccoli cultivars and its related species. Based on genetic similarity values, a UPGMA clustering dendrogram was successfully developed. At a similarity coefficient of 61% as a threshold level, the dendrogram was divided into three major groups (Fig. 3). Group I contained 27 cultivars, which then further divided into two subgroups at similarity coefficient of 73%. Subgroup I comprised 20 cultivars with the majority of the members are broccoli cultivars, except for one cauliflower cultivar (Violet Dream) and one kohlrabi cultivar (Worldcol). In this subgroup, two broccoli cultivars, Subaru and Grace, were found to be identical. On the other hand, subgroup II consisted of seven cultivars including two broccoli, three cauliflower and two kohlrabi cultivars. In which, two out of three cauliflower cultivars were undifferentiated. Group II represented four kohlrabi cultivars; meanwhile group III contained five cultivars including two broccoli, one cauliflower and two kohlrabi cultivars. Two broccoli cultivars which were member of group III, Tradition and Marathon, were also indistinguishable.

In addition, we also performed PCoA analysis based on genetic similarity value, which assigned 22 broccoli, 5 cauliflower and 9 kohlrabi cultivars into two groups (Fig. 4). The result exhibited that group I in PCoA resembles group I and II in UPGMA dendrogram, likewise group II in PCoA corresponds to group III in UPGMA dendrogram. Therefore, this PCoA analysis supports the grouping results developed by that of the UPGMA method.

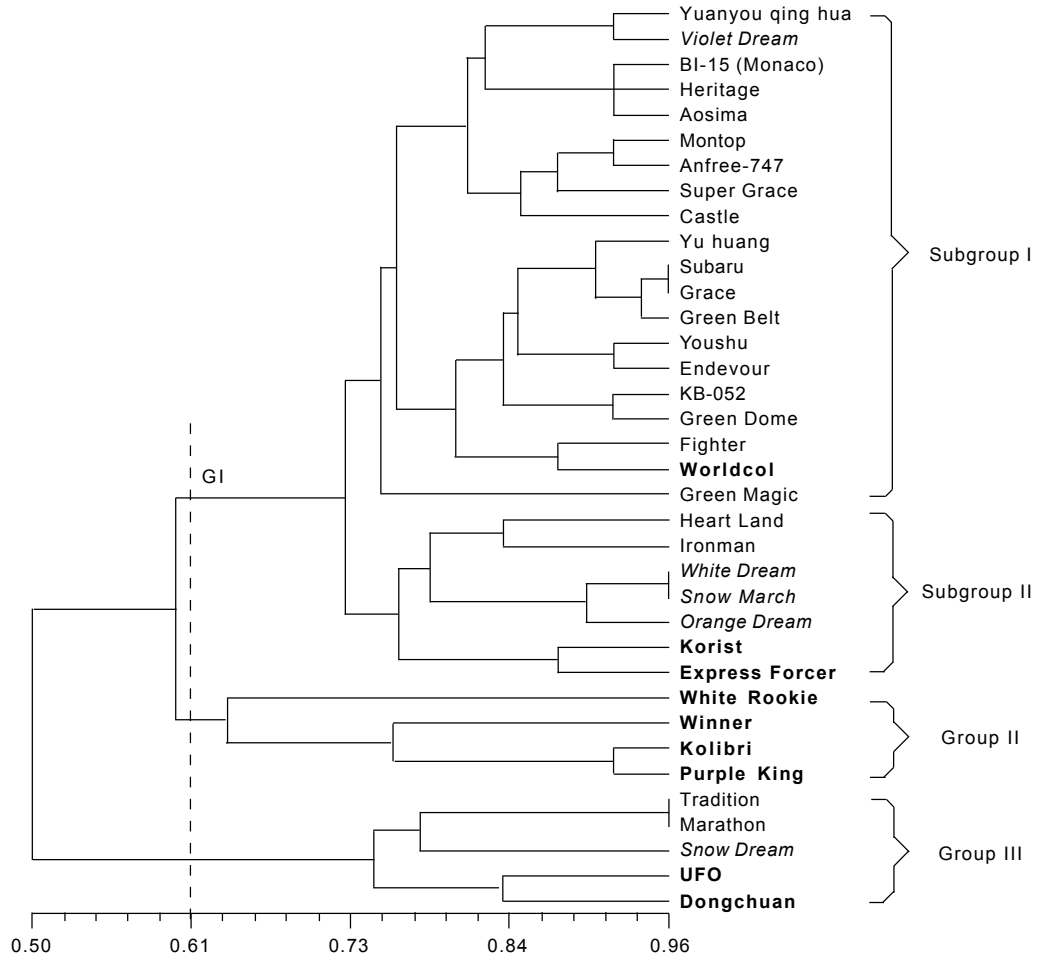

Fig. 3. Cluster dendrogram of thirty six *Brassica oleracea* cultivars developed based on genetic similarity coefficient using five newly designed EST-SSR markers. Broccoli cultivars were written in normal font, cauliflower cultivars in italic font, and kohlrabi cultivars in bold font.

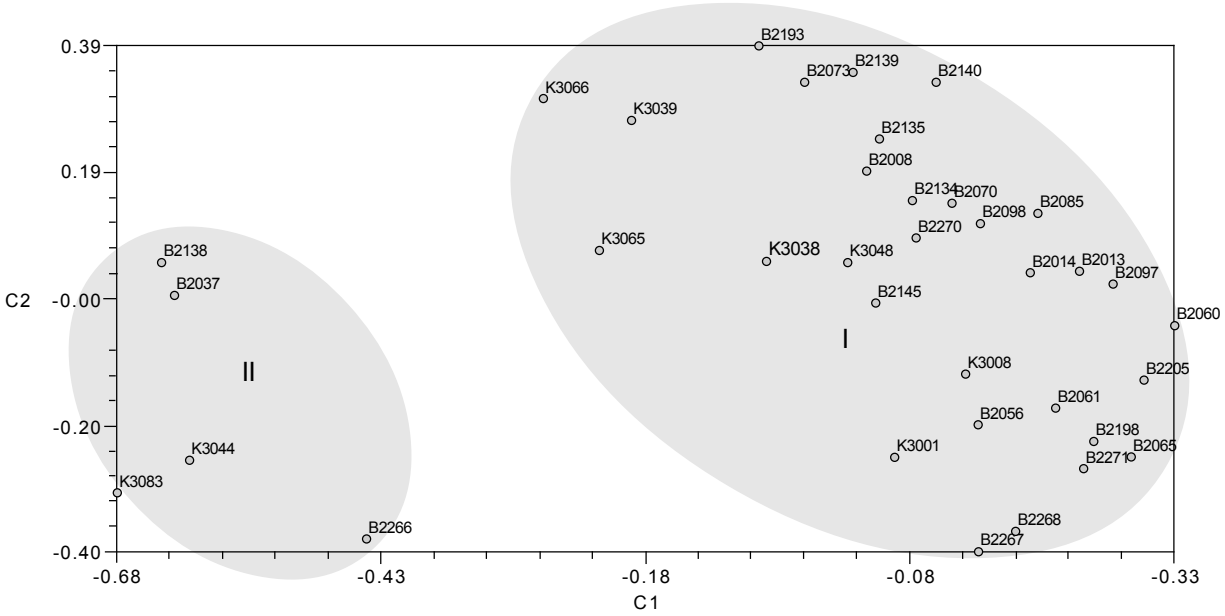

Fig. 4. PCoA analysis revealed relationships among thirty six broccoli cultivars and its related species established using five EST-SSR loci.

DISCUSSION

The availability of a large number of EST sequences has proven to be valuable as they provide important genetic information and plenty of gene-specific transcript sequences. As described by Ma *et al.* (2012), EST sequences facilitate designing DNA markers such as SSR markers through *in silico* mining. In present study, we report the characterization of 202 EST sequences of *B. oleracea* SSH-cDNA library of black rot infection in leaves obtained from public database with a total length of 68,925 kb as well as its usefulness in the development of EST-SSR markers. The development of an adequate number of species specific markers is crucial in order to allow many genetic analyses in broccoli and its related species. In our observation, we found that 5.94% of *B. oleracea* ESTs contained SSRs, which is in accordance with previous study conducted by Izzah *et al.* (2014). The observed frequency of SSRs in plant EST usually depends on several factors including the mining criteria and tools, EST datasets size and genome structure or composition (Varshney *et al.* 2005).

Furthermore, trinucleotide repeats were the most dominant repeat types obtained, which accounted for 50% among the identified SSRs. The abundance of trinucleotide repeats in SSRs derived from EST sequences were also found in other plant species such as citrus (45.1%), flax (76.8%), barley (56%) and jatropha (57.75%) (Thiel *et al.* 2003; Chen *et al.* 2006; Cloutier *et al.* 2009; Wen *et al*. 2010). The large number of trinucleotide repeats found in coding regions reflects the fact that this repeat type would not cause frame shift mutation that could silence the gene (Cloutier *et al.* 2009).

Since EST-SSR markers exhibited numerous advantages such as more feasible, low cost, fast and cost-effective approach in marker development over that of genomic SSRs, many scientists have begun to develop them in various plant species. Even though EST-SSR markers usually have lower polymorphism level compared to that of genomic SSRs, they showed higher level of transferability to related species due to the high level of conservation in the flanking SSRs sequences (Chen *et al*. 2015). Besides, EST-SSRs also showed more closely linked to underlying genes than genomic SSRs, which mean that the EST-SSRs mined herein may provide more information on genetic variation in *B. oleracea* genomes (Ma *et al.* 2012). In this study, we obtained that all of the designed primer were successfully amplified in all the cultivar DNA samples. The amplification results were higher than those of the earlier reports in several plant species that ranged from 60% to 90% (Varshney *et al.* 2005), which perhaps resulted from high quality of EST sequences used. Moreover, about 62.5% of new EST-SSR markers exhibited polymorphisms in a set of 22 broccoli cultivars and its related species. The polymorphism level gained in the present study was higher compared to those obtained from other plant species such as flax (40.7%), soybean (12.8%), and barley (35% and 42%) that were examined with EST-SSR markers (Thiel *et al.* 2003; Varshney *et al.* 2006; Hisano *et al.* 2007; Cloutier *et al.* 2009). The differences observed in the polymorphism levels might be influenced by the relatedness or the number of genotypes tested (Cloutier *et al.* 2009).

Polymorphism information content (PIC) that defined as a closely related diversity measure was also investigated. The average of PIC value found in the present study was 0.51, lower than that observed by Louarn *et al.* (2007) which was 0.64. The PIC value differences between the two studies may be because in this study we used EST-SSR markers that have more conserved sequences of gene-derived sequences, whereas the ones reported by Louarn *et al.* (2007) used genomic SSRs. Another possible explanation was the relatedness of the genotypes and the number of genotypes used were different in the two studies. However, identified polymorphisms in the present study directly sampled the variations in transcribed regions of the genome and reflected the diversity inside or adjacent to the genes, which make EST-SSR markers more attractive and informative than that of genomic SSRs (Hu *et al.* 2010). On the other hand, since these markers were developed from EST sequences of *B. oleracea* SSH-cDNA library of black rot infection in leaves, it would be great advantages for breeders working in broccoli and related species to support their breeding programs. This is because five polymorphic markers obtained in this study can be utilized by breeders for discriminating resistant and susceptible cultivars in broccoli. With the availability of these EST-SSR markers specific to black rot resistance, the selection process can be improved. This benefit can be exploited by breeders to accelerate breeding procedure for developing black rot resistant genetic stocks/lines in broccoli and its related species. Previous study also suggested the benefits of molecular markers to distinguish between resistant and susceptible varieties. For instance, Sharma *et al*. (2016) found three markers (At1g70610, Na14-G02, and At1g71865)

that can differentiate resistant and susceptible bulks against black rot disease in *Brassica carinata*.

The polymorphic EST-SSR markers developed in this study had successfully distinguished 30 out of 36 cultivars in cluster analysis. Meanwhile, six cultivars including four broccoli and two cauliflower cultivars were remain undifferentiated. The failure to differentiate those broccoli and cauliflower cultivars could be caused by their genetic close relatedness, although they come from different seed companies (Louarn *et al*. 2007). Therefore, addition of more polymorphic EST-SSR markers can be one of the solutions to improve the separation of the cultivars. Nevertheless, the novel EST-SSR markers designed in the present study could separate broccoli cultivars into two different groups, group I and group III. In which, 20 out of 22 broccoli cultivars were clustered together in group I. This result indicates that the broccoli cultivars which originated from different seed companies demonstrated close relatedness, which may be because those seed companies shared their breeding materials or they used common elite lines with different names (Izzah *et al*. 2013; Zhao *et al*. 2014). In general, the clustering of 22 broccoli cultivars did not reflect their main phenotypic characters. All of broccoli cultivars that placed in group I and group III showed different stages of maturity and various types of bead size and head shape. This finding is similar to the previous study reported by Izzah *et al*. (2013), which described that the grouping of all tested broccoli cultivars did not coincide with their phenotypic characters. However, we found two broccoli cultivars (Tradition and Marathon) clustered separately in group III. This indicates that these two cultivars are unique compared to the other cultivars, although they have similar phenotypic characters.

The grouping of cauliflower cultivars exhibited that most of them were clustered in group I, but located in different subgroups. Three cauliflower cultivars that clustered in subgroup 2 show the same period of maturity (medium) and have high domed shape. Meanwhile, cultivar Violet Dream that clustered in subgroup 1 has distinct phenotypic characters as follows: violet curd colour, early maturity and early bolting type. Interestingly, cultivar Snow Dream, which has same phenotypic characters with those of the three cultivars in subgroup 2, was placed in group III. In the case of kohlrabi cultivars, four cultivars clustered together in group II, and most of them exhibited early maturity and flat head shape. Meanwhile, the other five cultivars were scattered in group I and group III. Among nine kohlrabi cultivars, one cultivar (Worldcol) clustered separately in group I/subgroup 1. This outcome may correspond to character of very long field holding ability that only possessed by cultivar Worldcol. Overall, based on cluster analysis, we found five unique cultivars including two broccoli cultivars (Tradition and Marathon), two cauliflower cultivars (Violet Dream and Snow Dream), and one kohlrabi cultivar (Worldcol). All of these cultivars can be selected as a good candidate for parental lines in the future breeding program in order to maximize heterosis expression in their progenies.

CONCLUSION

The result presented in this study demonstrated the usefulness of EST sequences derived from public database for designing primer. A total of five polymorphic EST-SSR markers were generated. Cluster analysis using these polymorphic markers successfully divided 22 broccoli cultivars and its related species into three major groups. Group I dominated by broccoli cultivars, whereas group II only contained kohlrabi cultivars. The third group consisted of broccoli, cauliflower and kohlrabi cultivars. The grouping by UPGMA dendrogram also exhibited five unique cultivars comprising of two broccoli cultivars, two cauliflower cultivars and one kohlrabi cultivar that can be used as valuable breeding materials. Furthermore, five polymorphic EST-SSR markers obtained here would be useful for differentiating resistant and susceptible cultivars against black rot disease, because these markers were developed from *B. oleracea* SSH-cDNA library of black rot infection in leaves. The outcome of this study should be useful in supporting breeding program of broccoli and its related species.

ACKNOWLEDGEMENTS

This work was supported by the Technology Development Program (No. 309008-05) for Agriculture and Forestry, Ministry of Food, Agriculture, Forestry and Fisheries, Republic of Korea. We would like to thank Dr. Kyounggu Ahn from the Joeun Seed Company, who has provided all the plant materials used in present study.

REFERENCES

Ali, G.M., S. Yasumoto and M. Seki-Katsuta. 2007. Assessment of genetic diversity in sesame (*Sesamum indicum* L.) detected by amplified fragment length polymorphism markers. Electronic J. Biotechnol. 10: 1.

Allen, G.C., M.A. Flores-Vergara, S. Krasynanski, S. Kumar and W.F. Thompson. 2006. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nat Prot. 1: 2320–2325.

Chen, C., P. Zhou, Y.A. Choi, S. Huang and F.G. Gmitter Jr. 2006. Mining and characterizing microsatellites from citrus ESTs. Theor. Appl. Genet. 112: 1248–1257.

Chen, H., L. Liu, L. Wang, S. Wang, P. Somta and X. Cheng. 2015. Development and validation of EST-SSR markers from the transcriptome of Adzuki bean (*Vigna angularis*). PlosOne. 10(7): e0131939. Doi: 10.1371/journal.pone.0131939.

Cloutier, S., Z. Niu, R. Datla and S. Duguid. 2009. Development and analysis of EST-SSRs for flax (*Linum usitatissimum* L.). Theor Appl Genet. 119: 53–63.

Hisano, H., S. Sato, S. Isobe, S. Sasamoto, T. Wada, A. Matsuno, T. Fujishiro, M. Yamada, S. Nakayama, Y. Nakamura, S. Watanabe, K. Harada and S. Tabata. 2007. Characterization of the soybean genome using EST-derived microsatellite markers. DNA Res. 14: 271–281.

Hu, J-B., X-Y. Zhou and J-W. Li. 2010. Development of novel EST-SSR markers for cucumber (*Cucumis sativus*) and their transferability to related species. Scientia Horticulturae 125: 534–538.

Huang, X. and A. Madan. 1999. CAP3: a DNA sequence assembly program. Genome Res. 9: 868–877.

Izzah, N.K., J. Lee, S. Perumal, J.Y. Park, K. Ahn, D. Fu, G-B. Kim, Y-W. Nam and T-J. Yang. 2013. Microsatellite-based analysis of genetic diversity in 91 commercial *Brassica oleracea* L. cultivars belonging to six varietal groups. Genet. Resour. Crop Evol. 60: 1967–1986.

Izzah, N.K., J. Lee, M. Jayakodi, S. Perumal, M. Jin, B-S. Park, K. Ahn and T-J. Yang. 2014. Transcriptome sequencing of two parental lines of cabbage (*Brassica oleracea* L. var. *capitata* L.) and construction of an EST-based genetic map. BMC Genomics 15: 149.

Kalia, R.K., M.K. Rai, S. Kalia, R. Singh and A.K. Dhawan. 2011. Microsatellite markers: an overview of the recent progress in plants. Euphytica 177: 309–334.

Kumari, M., A. Grover, P.V. Yadav, M. Arif and Z. Ahmed. 2013. Development of EST-SSR markers through data mining and their use for genetic diversity study in Indian accession of *Jatropha curcas* L.: a potential energy crop. Genes & Genomics 35(5): 661–670.

Kumpatla, S.P. and S. Mukhopadhyay. 2005. Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. Genome. 48: 985–998.

Lee, J., N.K. Izzah, M. Jayakodi, S. Perumal, H.J. Joh, H.J. Lee, S-C. Lee, J.Y. Park, K-W. Yang, I-S. Nou, J. Seo, J. Yoo, Y. Suh, K. Ahn, J.H. Lee, G.J. Choi, Y. Yu, H. Kim and T-J. Yang. 2015. Genome-wide SNP identification and QTL mapping for black rot resistance in cabbage. BMC Plant Biol. 15: 32.

Liu, K. and V. Muse. 2005. PowerMarker: integrated analysis environment for genetic marker data. Bioinformatics 21: 2128–2129.

Louarn, S., A.M. Torp, I.B. Holme, S.B. Andersen and B.D. Jensen. 2007. Database derived microsatellite markers (SSRs) for cultivar differentiation in *Brassica oleracea*. Genet. Resour. Crop Evol. 54: 1717–1725.

Lu, X., L. Liu, Y. Gong, L. Zhao, X. Song and X. Zhu. 2009. Cultivar identification and genetic diversity analysis of broccoli and its related species with RAPD and ISSR markers. Scientia Horticulturae 122: 645–648.

Ma, J-Q., C-L. Ma, M-Z. Yao, J-Q. Jin, Z-L. Wang, X-C. Wang and L. Chen. 2012. Microsatellite markers from tea plant expressed sequence tags (ESTs) and their applicability for cross-species/genera amplification and genetic mapping. Scientia Horticulturae 134: 167–175.

Meenu, G., A. Vikram and N. Bharat. 2013. Black rot - a devastating disease of Crucifers: A reviews. Agric. Rev. 34(4): 269–278.

Rohlf, F.J. 2000. NTSYS-PC, Numerical Taxonomy System for the PC, ExeterSoftware, Ver. 2.1. Applied Biostatistics Inc, Setauket.

Sharma, B.B., P. Kalia, D.K. Yadava, D. Singh and T.R. Sharma. 2016. Genetics and molecular mapping of black rot resistance locus *Xca1bc* on chromosome B-7 in Ethiopian mustard (*Brassica carinata* A. Braun). PlosOne. 11(3): e0152290. doi: 10.1371/journal.pone.0152290.

Shirasawa, K., M. Oyama, H. Hirakawa, S. Sato, S. Tabata, T. Fujioka, C. Kimizuka-Takagi, S. Sasamoto, A. Watanabe, M. Kato, Y. Kishida, M. Kohara, C. Takahashi, H. Tsuruoka, T Wada, T. Sakai and S. Isobe. 2011. An EST-SSR linkage map of *Raphanus sativus* and comparative genomics of the Brassicaceae. DNA Res. 18: 221–232.

Temnykh, S., G. DeClerck, A. Lukashova, L. Lipovich, S. Cartinhour and S. McCouch. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res. 11: 1441–1452.

Thiel, T., W. Michalek, R.K. Varshney and A. Graner. 2003. Exploiting EST databases for the development of cDNA derived microsatellite markers in barley (*Hordeum vulgare* L.). Theor Appl Genet. 106: 411–422.

Varshney, R.K., A. Graner and M.E. Sorrells. 2005. Genic microsatellite markers in plants: features and applications. Trends Biotechnol. 23: 48–55.

Varshney, R.K., I. Grosse, U. Hahnel, R. Siefken, M. Prasad, N. Stein, P. Langridge, L. Altschmied and A. Graner. 2006. Genetic mapping and BAC assignment of EST-derived SSR markers shows non-uniform distribution of genes in the barley genome. Theor Appl Genet. 113: 239–250.

Wang, Z., J. Li, Z. Luo, L. Huang, X. Chen, B. Fang, Y. Li, J. Chen and X. Zhang. 2011. Characterization and development of EST-derived SSR markers in cultivated sweet potato (*Ipomoea batatas*). BMC Plant Biology. 11: 139.

Wen, M., H. Wang, Z. Xia, M. Zhou, C. Lu and W. Wang. 2010. Development of EST-SSR and genomic-SSR markers to assess genetic diversity in *Jatropha curcas* L. BMC Research Notes. 3: 42.

Yi, G., J.M. Lee, S. Lee, D. Choi and B-D. Kim. 2006. Exploitation of pepper EST–SSRs and an SSR-based linkage map. Theor Appl Genet. 114:113-–130.

Zhao, Z., H. Gu, X. Sheng, H. Yu, J. Wang, J. Zhao and J. Cao. 2014. Genetic diversity and relationships among loose-curd cauliflower and related varieties as revealed by microsatellite markers. Scientia Horticulturae. 166:105–110.

Zhou, Q., T. Chen, Y. Wang and Z. Liu. 2014. The development of 204 novel EST-SSRs and their use for genetic diversity analyses in cultivated alfalfa. Biochemical Systematics and Ecology. 57: 227–230.