Full Length Research Paper

Genome shotgun sequencing and development of microsatellite markers for gerbera (*Gerbera hybrida* H.) by 454 GS-FLX

Kyoung-In Seo[1], Gi-An Lee[1], Sang-Kun Park[2], Mun-Sup Yoon[1], Kyung-Ho Ma[1], Jung-Ro Lee[1], Yu-Mi Choi[1], Yeon-ju Jung[1] and Myung-Chul Lee[1]*

[1]National Agrobiodiversity Center, National Academy of Agricultural Science, RDA, 88-20, Seodun-Dong, Suwon, Gyunggi-do, 441-707, Korea.
[2]Floriculture Research Division, National Institute of Horticultural and Herbal Science, RDA, Suwon, Gyunggi-do, 440-706, Korea.

The objective of this research was to develop and characterize microsatellite markers for gerbera. We used shotgun sequencing with Roche 454 GS-FLX Titanium technology to identify microsatellite loci in gerbera genomic DNA (*Gerbera hybrida*). The total length of non-redundant sequences obtained was 22,527,019 bp, which consisted of 3,085 contigs and 28,249 singletons. We assembled 61,958 reads into 3,085 contigs, of which 114 (3.70%) contained microsatellite repeats. The average G+C content was 39.3%. Functional annotation to known sequences yielded 14.7% unigenes in the 'Raon' cultivar. Analysis of the gerbera genome DNA ('Raon') general library showed that sequences of (AT), (AG), (AAG) and (AAT) repeats appeared most often, whereas (AC), (AAC) and (ACC) were the least frequent. Primer pairs were designed for 80 loci. Only eight primer pairs produced reproducible polymorphic bands in the 28 gerbera accessions analyzed. A total of 30 alleles were identified from the eight polymorphic SSR loci, with two to eight alleles per locus (average level of 3.75). These markers will be useful for investigating genetic diversity and differentiation in gerbera.

Key words: Genetic diversity, genomics, microsatellite isolation, pyrosequencing, SSRs.

INTRODUCTION

Gerbera species, belonging to the family Asteraceae, are among the most popular decorative plants worldwide, with uses as garden plants, potted plants and cut flowers. Gerbera plants generally show variation in their stalks, and through hybrid breeding (they are polyploid) thin, soft and cut flower varieties have been produced (Kanwar and Kumar, 2008; Yu et al., 1999). Most of the gerbera cultivated in Korea have large flowers that are used as cut flowers. Various cut flower gerbera cultivars have different flower colors. In addition, year-round production is possible and is a common practice. Evaluations of the genetic diversity of gerbera have been conducted using random amplified polymorphic DNA (RAPD) and

expressed sequence tag-simple sequence repeats (EST-SSR) (Mata et al., 2009; Gong and Deng, 2010). In addition, Bhatia et al. (2009) studied the effects of explant source on the genetic integrity of 45 *in vitro*-raised gerbera plants (*Gerbera jamesonii* Bolus) using 32 ISSR markers. However, microsatellite (SSRs) markers were not developed and tested for these gerbera genetic studies. Simple sequence repeats (SSR) or micro-satellites have been used to examine genotypic variation and for breeding research (Eujayl et al., 2001; Ruiz et al., 2000). Furthermore, these markers are widely used in cultivar fingerprinting and molecular mapping. When compared with other types of molecular markers, SSR markers are abundant, co-dominant and have high reproducibility. In the past, microsatellite markers were developed using the enrichment method and hybridization probes. However, this method had a disadvantage: it could not detect the correct repeat sequence in the

*Corresponding author. E-mail: mclee@rda.go.kr. Tel: 82-31-299-1887. Fax: 82-31-294-6029.

Table 1. Gerbera accessions analyzed using SSR markers and characteristics of gerbera flower types.

S/N	Accession name	Flower type
1	Marathon	Semi-double
2	Venturi	Double
3	Samba	Single
4	Angi	Semi-double
5	Liz	Semi-double
6	Kopabo	Single
7	Amado	Semi-double
8	Mirage	Double
9	Goldfinger	Double
10	Kabana	Double
11	Albert	Double
12	Diablo	Semi-double
13	Esperanza	Double
14	Sunset	Double
15	Rosamatt	Double
16	Ma	Semi-double
17	Duella	Semi-double
18	Essence	Double
19	Salina	Double
20	Palazo	Double
21	Escada	Semi-double
22	Stephane	Semi-double
23	Raon	Double
24	Misty Red	Double
25	Royal	Semi-double
26	Mandorin	Semi-double
27	Jinno	Semi-double
28	Dino	Double

generation sequencing (Sexton et al., 2010).

Many studies have used 454 pyrosequencing to analyze genomic DNA (Sato et al., 2010), polymerase chain reaction (PCR) amplicons (Bonnal et al., 2010), BAC DNA (Wicker et al., 2006) and cDNA (Yang et al., 2010) samples. Tangphatsornruang et al. (2009) developed polymorphic SSR markers for genotyping mungbean using genome shotgun sequencing. They reported that shotgun sequences were useful for marker development, for determining phylogenetic relationships and for gene discovery. The present study examined microsatellite sequences in gerbera genomic data and evaluated genetic relationships between gerbera cultivars. In this paper, we reported the development of micro-satellites and the characterization of gene content in gerbera (*Gerbera hybrida*) using cost-effective 454-pyrosequencing.

MATERIALS AND METHODS

Plant material and DNA extraction

PCR and genotyping was performed on 28 gerbera (*Gerbera hybrida* H.) accessions collected from the National Horticultural Research Institute (NHRI), Rural Development Administration (RDA) of Korea (Table 1). The 'Raon' and 'Misty Red' cultivars were bred at the NHRI of Korea and the rest of the cultivars were bred abroad. The scoring of flower type was based on the number of petals, such as single, double and semi-double. Among 28 accessions, only 2 accessions have single type flower and 12 other accessions were semi-double type and 14 accessions have double type.

Total genomic DNA was extracted from young fresh leaves by the modified cetyltrimethylammonium bromide (CTAB) method described by Dellaporta et al. (1983). A total of 8 µg of 'Raon' (double petal type) DNA was used to perform shotgun sequencing using 454 GS-FLX Titanium pyrosequencing (Roche Applied Science, Mannheim, Germany).

Shotgun genomic sequencing

A small/medium plate (1/8 PicoTiterPlate; PTP) was used for 454 GS-FLX Titanium pyrosequencing for samples processed by the shotgun method according to the manufacturer's instructions. In short, the samples prepared using the shotgun method was fractionated into smaller pieces (300 to 800 bp) by nebulization. Subsequently, the fragment length distribution of the library was assessed using a gerbera DNA 7500 LabChip (Agilent Technologies, Waldbronn, Germany) to ensure that the input material was adequate. The ends of the sheared gerbera genome DNA fragments were polished by the action of T4 DNA polymerase and T4 polynucleotide kinase. Short adaptors were then ligated to the ends of each sample DNA fragment. These adaptors provide priming regions to support both amplification and nucleotide sequencing. Finally, the quality of the library was assessed with a Bioanalyzer RNA 6000 Pico LabChip (Agilent Technologies).

Emulsion titration was performed with different amounts of input ssDNA. Medium-scale emPCR was conducted based on the titrated ssDNA amount. Enriched bead samples were then counted using a Z1 Coulter Counter (Beckman Coulter, Brea, CA, USA) to calculate the percent enrichment.

Samples prepared using the above procedure were loaded,

genome because a probe that facilitated hybridization was not available.

Lagercrantz et al. (1993) demonstrated that (AT)/(TA) and (CG)/(GC) probes more easily construct comple-mentary structures than do other kinds of probes. In addition, the method was costly, labor intensive during the cloning process, and produced many redundant clones. The recent introduction of next-generation sequencing technologies, such as the 454 Genome Sequencer (GS)-FLX (Roche) which uses massively parallel pyrosequencing, may allow for more effective production of microsatellites. The GS-FLX by 454 Life Sciences depends on an oil-emulsion based PCR followed by massively parallel and individual pyro-sequencing of the clonally amplified beads in a PicoTiterPlate (Pettersson et al., 2009). At present, 454 technology offers the longest read length by far, ~400 bp on the GS-FLX Titanium platform. Recently, in *Khaya senegalensis*, a microsatellite marker that can charac-terize genetic variation was developed using next-

along with other reagents, into the wells of a 1/8 PTP device, and sequencing was performed with a GS FLX Titanium Reagents XLR70, a 70 x 75 mm Titanium PTP device and a Genome Sequencer FLX Instrument (454 Life Science, Roche Applied Science). Sequencing data were obtained after a 9-h run of the GS-FLX.

After sequencing, sequence assembly was performed using the GS De Novo Assembler software to produce the contigs and singletons. All sequence alignments were conducted using the Basic Local Alignment Search Tool (BLAST; database: NT) (Altschul et al., 1997).

Selection of SSRs and primer design

Identification of SSR motifs within contig sequences and primer design were conducted using the ARGOS program (Kim, 2004). The contigs were screened for mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide repeats. The minimum number of repeats was set to four. From these contigs, we designed 80 primer pairs to amplify microsatellite regions. The primer design included the following specifications: (1) melting temperature (Tm) between 54 and 64°C; (2) GC content above 30%; (3) amplicon size of 150 to 300 bp. For each locus, forward primers were synthesized to incorporate a 5' fluorescent label, M13 (-21)-Tag TGTAAAACGACGGCCAGT (Messing, 1983).

Polymorphism analysis

The M13F-tail (5′-TGTAAAACGACGGCCAGT-3′) PCR method of Schuelke (2000) was used to measure PCR product size, as described by Lee et al. (2008). PCR was performed in a 20 μl reaction mixture containing 2 μl of template genome DNA (20 ng/μl), 0.2 μl of locus-specific primer (10 pmol/μl), 0.4 μl of M13 universal primer (10 pmol/μl), 0.6 μl normal reverse primer, 2.0 μl of 10X PCR buffer (SolGent, Daejeon, Korea), 1.6 μl of dNTP (2.5 mM) and 0.2 μl h-Taq-polymerase (5 U/μl; SolGent). The 5'-M13 sequence was attached to a forward primer to incorporate a florescent dye [blue (FAM), green (NED) or yellow (HEX)] into the PCR product. PCR was performed as follows: initial denaturation at 94°C (3 min), then 30 cycles each of 94°C (30 s), 55°C (45 s) and 72°C (1 min), followed by 10 cycles of 94°C (30 s), 53°C (45 s) and 72°C (1 min), and a final extension at 72°C for 10 min (PTC-100 Thermocycler, MJ Research, Waltham, MA, USA). All denatured products were run on an ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems, Carlsbad, CA, USA) using the molecular size standard (35 to 500 bp) GeneScan 500 ROX (6-carboxy-rhodamine). Genotypes were obtained using GeneMapper v4.0 software (Applied Biosystems).

Data analysis

PCR genotyping results were converted to binary form (0 = absence; 1 = presence), and similarity coefficients were calculated among all 28 accessions considered in this study with the program NTSYS, version 2.1 (Exeter Software, Setauket, NY, USA), using SIMQUAL with the Dice coefficient (Dice, 1945). The data were subjected to unweighted pair group method with arithmetic mean (UPGMA) cluster analysis using NTSYS 2.1. The number of distinct genotypes and genetic diversity were computed using GenoType/GenoDive software (Meirmans and Van Tienderen, 2004).

Sequence annotation and analysis of gene content

Functional annotation and major functional categories (biological process, molecular function and cellular component) were analyzed

using the BLAST2GO software suite v2.3.1 (www.blast2go.de) (Conesa and Götz, 2008). Homology searches (BLASTX and BLASTN) were performed remotely on the NCBI server through a query-friendly version of BLAST (QBLAST), following a sequential strategy.

RESULTS

Sequence analysis of the gerbera genome

The status yield of sequencing runs and statistics for 'Raon' are summarized in Table 2. We also provided sequence data for all contigs of 'Raon' as Supplementary File 1. The 454 GS-FLX Titanium sequencing run yielded 22.5 Mbp of a genome DNA library from 'Raon' (*G. hybrida*), including 22,527,019 reads, with an average read length of 363.6 bp and 114 contigs (3.69%) containing microsatellite repeats. The 'Raon' genome DNA was assembled into 3,085 contigs and 28,249 singletons. After the assembly (3,085 contigs + 28,249 singletons), the contigs had an average length of 344.03 bp and an average depth of coverage of 10.92 reads per base pair. A total of 430 large contigs (≥500 bp), with an average length of 1,046 bp, was detected for 'Raon'. The longest contig was 25,066 bp.

To better understand the general mechanisms working in the gerbera, we performed the gene ontology (GO) analysis and carried out all aligned sequences (3,085 contigs and 28,249 singletons). Functional annotation was analyzed using BLAST2GO (Conesa and Götz, 2008), which led to consistent gene annotations and the assignment of gene names, gene products and gene ontology (GO) numbers. In total, 452 unigenes were assembled, and assembled contig sequences were available. All unigenes were investigated for their association with GO terms from level 4 biological processes, molecular function and cellular components; the number of associations was 137 (30.3%), 85 (18.8%) and 56 (12.4%), respectively. Eighty-five assignments were made to molecular function ontology, with a large proportion of these functioning in DNA binding (32.7%) and cation binding (27.3%) (Figure 1). Under biological process ontology, 137 assignments were made, with a large proportion of the assignments falling into cellular macromolecule metabolic process (24%) and nucleobase, nucleoside, nucleotide and nucleic acid metabolic process (23%) categories. The top 10 contigs by number of assembled reads were aligned to the NCBI database using BLASTN (Table 3). Four had significant (e-value ≤1 x 10^{-3} and a bit score >40) hits to different genes, while three aligned to *Lactuca sativa* cultivar Salinas chloroplast complete genome DNA and three others aligned to *Guizotia abyssinica* chloroplast complete genome DNA.

Characteristics of the genome SSR

SSR motifs found in 'Raon' are summarized in   Figure 2.

Table 2. Sequencing results by GS FLX Titanium.

Total bases (bp)	22,527,019
Significant BLAST hits[Z]	152
G+C content (%)	39.3
Length Average (bp)	363.58
Total Contig length (bp)	1,061,348
Number of contigs	3,085
Mean depth of contig	10.92
Number of singletons	28,249
Percentage of reads containing microsatellites	3.69%
Sequences with amplifiable microsatellite sequences	0.29%

[Z]Significant BLAST hit criteria was an e-value $\leq 1 \times 10^{-3}$ and a bit score>40.

Of the 3,085 unique contigs, 114 (3.70%) contained microsatellite sequences. Among the 114 microsatellite regions, we found dinucleotide, trinucleotide and pentanucleotide repeat motifs. The frequency of the appearance of these SSRs was estimated to be one SSR for every 9.3 kb in the 1.06 Mbp sequences from the *Gerbera* genome. The repeat number for dinucleotides (62.3%) was higher than that for trinucleotides (36.8%). In addition, a pentamer repeat was found in one contig (0.9%). The number of repeats ranged from 4 to 33. Among the SSRs, AT/TA repeat motifs were the most common (28.9%) followed by AG/GA repeat motifs (21.1%). The most frequent trinucleotide motif was AAG/AGA/GAA (13.2%), which was closely followed by the AAT/ATA/TAA motif (7.9%).
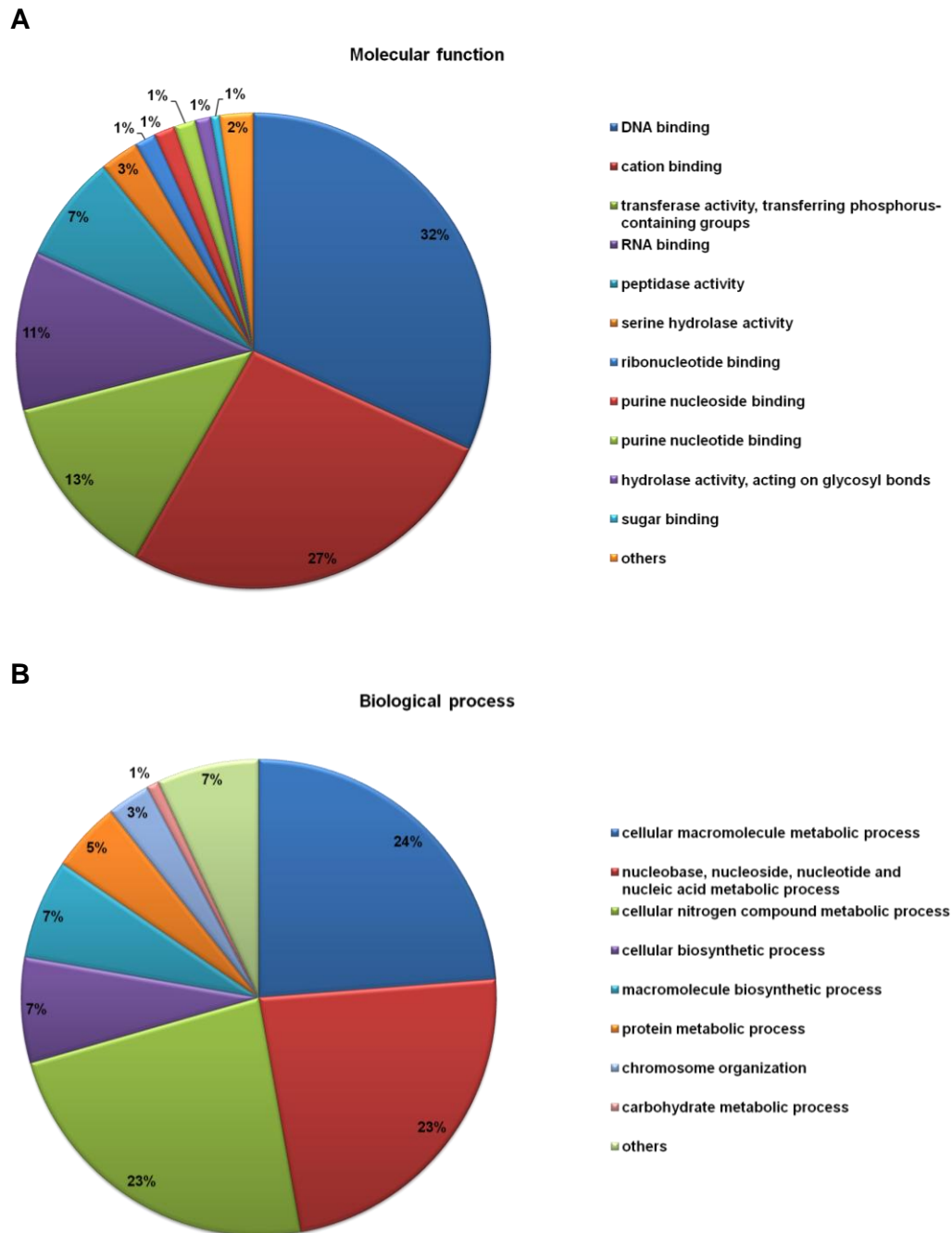
Development of SSRs markers

After screening all the contigs (n = 3,085) for SSR regions with the ARGOS PROGRAM (Kim, 2004), we identified a total of 114 SSR regions. A random set of 80 primer sets was designed which included SSR regions; these primers were used to amplify gerbera genomic regions. There were six perfect repeats having no interruptions, one imperfect repeat and one compound repeat sequence. Four of the loci: GB-GJ-002, GB-GJ-049, GB-GJ-059 and GB-GJ-063, which had $(AAT)_4$, $(AT)_4$, $(ACA)_6$, $(AG)_7$, $(ATC)_4$, and $(CAT)_5$ repeats, respectively, were labeled with FAM (blue). Two others, GB-GJ-011 and GB-GJ-074, with $(CAC)_5$ and $(CAT)_5$ repeats, respectively, were labeled with HEX (yellow). The last loci, GB-GJ-021 and GB-GJ-077, which had $(AAAG)_4$ and $(TGA)_6$ repeats, respectively, were labeled with NED (green). GB-GJ-049 had the widest   allele size

range of the eight loci (160 to 317 bp) (Table 4). When used with 28 accessions, eight primer pairs managed to amplify 30 distinct alleles, with an average of 3.75 alleles per locus.

Genetic diversity based on SSR data

Genetic diversity among *G. hybrida* accessions was evaluated using an UPGMA model (Figure 3). Cluster analysis separated the three groups: Groups-I, II and III. Group-I accommodated the highest number of accessions (16 accessions, 57.1%) followed by Group-II (nine accessions, 32.1%). Group-III contained the lowest number of accessions (three accessions, 10.7%). As shown in Figure 3, most of the semi-double type accessions were clustered in Group-I. Group-II consisted of nine accessions, which were mainly double type; Group-III consisted of three accessions: one single type, one double type, and one semi-double type. Coefficients of similarity ranged from 0.68 to 0.96. Flower type did not differentiate clearly among gerbera accessions. The 'Misty Red' and 'Raon' accessions, which were uniquely bred in Korea (Choi et al., 2001), belonged to Groups-I and II, respectively. Our analyses of genetic distances (NTSYS) and calculations of clonal diversity (GenoType/ GenoDive) divided the *G. hybrida* accessions in Group-I into 14 genotypes. The Groups-II and III populations were divided into eight and three genotypes, respectively. This is because SSR markers detected multilocus genotypes within the groups. Calculations of clonal diversity (Genotype/Genodive) resulted in a Nei's diversity index (corrected for sample size) of 0.98 for Group-I (Table 5). For Groups-II and III, clonal diversity was 0.97 and 1.00, respectively, as nine and three genotypes were identified,

A

Molecular function

- DNA binding
- cation binding
- transferase activity, transferring phosphorus-containing groups
- RNA binding
- peptidase activity
- serine hydrolase activity
- ribonucleotide binding
- purine nucleoside binding
- purine nucleotide binding
- hydrolase activity, acting on glycosyl bonds
- sugar binding
- others

B

Biological process

- cellular macromolecule metabolic process
- nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
- cellular nitrogen compound metabolic process
- cellular biosynthetic process
- macromolecule biosynthetic process
- protein metabolic process
- chromosome organization
- carbohydrate metabolic process
- others

Figure 1. Gene ontology classification for gerbera according to molecular function (A) and biological process (B) using BLAST2GO with E-6 cutoff.

respectively. Total diversity over all accessions was 0.99 and average diversity within populations was 0.98. The fraction of diversity among populations, corrected G'_{ST}, was 0.01.

DISCUSSION

The 454 GS-FLX technology provided a very cost-

effective method for obtaining SSR sequences. This technology will accelerate research on less-studied gerbera plants. The 454 GS-FLX has been broadly used for transcriptome sequencing, whole genome sequencing, metagenomics and microbial diversity (http://454.com/applications/index.asp). As next-generation sequencing, 454 GS-FLX Titanium possesses more powerful sequencing capabilities, with a read length of 400 bp and higher sequencing accuracy. In this study, *G. hybrida*

Table 3. The top ten contigs determined by the number of assembled reads and their respective alignments.

Contig	BLASTN result	Contig Length (bp)	No. of assembled reads
1931	*Lactuca sativa* cultivar Salinas chloroplast, complete genome	25066	916
1954	*Guizotia abyssinica* chloroplast, complete genome	18240	460
2020	*Corokia cotoneaster* 26S ribosomal RNA gene, complete sequence	6606	319
28	*Guizotia abyssinica* chloroplast, complete genome	14762	306
5	*Guizotia abyssinica* chloroplast, complete genome	11466	253
56	*Trachelium caeruleum* chloroplast, complete genome	10443	230
3831	*Oryza punctata*, complete sequence	1296	226
1446	*Oryza australiensis*, complete sequence	5698	218
2	*Lactuca sativa* cultivar Salinas chloroplast, complete genome	6924	150
64	*Lactuca sativa* cultivar Salinas chloroplast, complete genome	7569	134
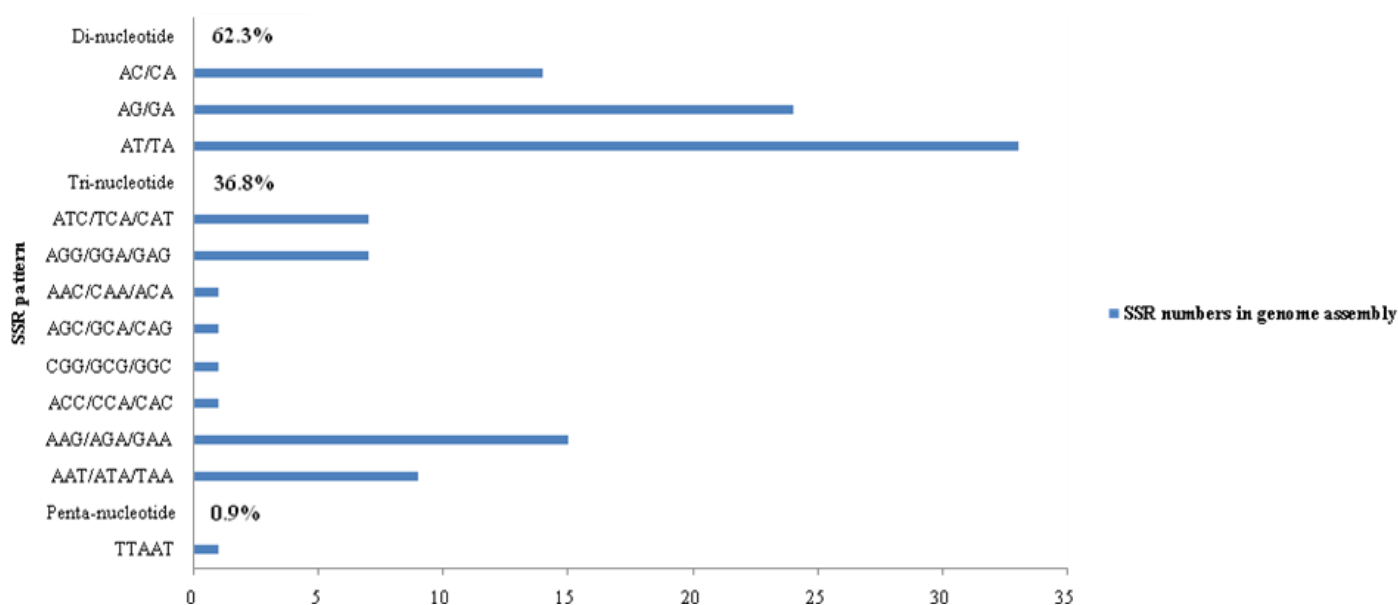
Figure 2. Distribution pattern of SSRs found in sequence of the 'Raon' cultivar.

genome DNA was sequenced using a 454 GS-FLX system, producing a total of 22,527,019 sequence reads before assembly (reads only). After assembly, 3,085 contig sequences were obtained. The GC content of all reads was 39.3%, which is within the range commonly observed for plant genomes [e.g., tomato (Zhu et al., 2008) and Arabidopsis (Arabidopsis Genome Initiative, 2000)]. Our sequencing runs yielded reads with an average length of 363.5 bp.

In the present study, 3,085 contigs (about 1.1 Mb) were used in an SSR search, which produced 114 (3.69%) SSR-containing sequences, including at least four dinucleotide, trinucleotide, tetranucleotide, penta-nucleotide and hexanucleotide repeats. This is a relatively low abundance of SSRs as compared to the numbers of ESTs containing SSRs found in an earlier report: *Citrus* spp. 6.09%, *Lactuca sativa* 4.90%, *Rosa* spp. 8.05% and

Prunus persica 4.99% (Kumpatla and Mukhopadhyay, 2005). The occurrence of SSRs in ESTs from the gerbera transcriptome was at the rate of one SSR per 5.16 kb (Gong and Deng, 2010), whereas in this study, using genomic sequences (represented by 1.1 Mbp of gerbera genome DNA), the frequency was one SSR for every 9.3 kb.

The proportion of AT/TA motifs in the SSRs generated from the gerbera genome was much higher than the proportion of (CA)/(TG) repeat motifs (Figure 2). This result is consistent with surveys of microsatellites in *Jatropha curcas* L. (Sato et al., 2010) and also with studies on barley (Becker and Heun, 1995), *Cannabis sativa* (Algharim and Almirall, 2003) and citrus (Chen et al., 2006). However, studies of *Brassica napus* (Uzunova and Ecke, 1999) and *Phaseolus vulgaris* (Benchimol et al., 2007) found greater abundances of (CA)/(TG) motifs.

Table 4. Characteristics of polymorphic microsatellite markers isolated for gerbera.

| Locus	GenBank accession number	Repeat motif	Repeat status	Primer sequence (5'→3')	Fluorescent dye	Tm [Z] (°C)	Expected Product size	Gerbera size range (bp)	# Alleles[W]
GB-GJ-002	JF343527	$(AAT)_4$	Perfect	F: TCAGTCTTCTTTGCCGTTTC	FAM	58	188	203-206	2
				R: TGATCCACGATTTCCCAG					
GB-GJ-011	JF343528	$(CAC)_5$	Perfect	F: GTGCATGGCTGTGTTCAA	HEX	58	255	234-273	4
				R: CCCAGGACAAAGTAGCCC					
GB-GJ-021	JF343529	$(AAAG)_4$	Perfect	F: GCCTCCGAATTTGCTTTT	NED	58	283	299-303	2
				R: TTCAAGCATGAGGGCAAG					
GB-GJ-049	JF343530	$(AT)_4$, $(ACA)_6$	Compound	F: AATTGTCCTCGACTCCGTT	FAM	58	296	160-317	8
				R: ATCCATGTAGCCAACCCC					
GB-GJ-059	JF343531	$(AG)_7$	Perfect	F: TGTCACTGTGTGAGAGTCGC	FAM	58	189	182-211	4
				R: GGATGACGATGACGAGGA					
GB-GJ-063	JF343532	$(ATC)_4$, $(CAT)_5$	Imperfect	F: TTTCCCCCATCATCCTTT	FAM	58	192	194-235	4
				R: GGATGACGATGACGAGGA					
GB-GJ-074	JF343533	$(CAT)_5$	Perfect	F: TCTCAAATTGCCTTCCGA	HEX	58	283	296-299	2
				R: GGTCTATCAAGGGGCCAG					
GB-GJ-077	JF343534	$(TGA)_6$	Perfect	F: GAGCAAGCCATCATCACC	NED	58	197	195-218	4
				R: CTTGCTTTCCTGCTCCCT					

[Z]Tm, annealing temperature; #Alleles[W], the total number of observed alleles among the 28 accessions.

Malausa et al. (2011) reported that enriched libraries are able to design primers about three times more than shotgun library. But, the genomic library enriched method have the disadvantage of the difficulty encountered in hybridization with AT and GC probe, which makes it to easily construct the self-complementary structure than the other kinds of repeat types of DNA sequences (Ashkenazi et al., 2001; Lagercrantz et al., 1993). Malausa et al. (2011) analyzed *Apis mellifera* using enriched and shotgun libraries; the percentage of microsatellite motifs of AT nucleo-tides was 8 and 39%, respectively. These results were similar to that detected in our gerbera shotgun library.

The importance of microsatellites as genetic markers and their potential use in various applications, such as diversity analysis or genetic mapping for crop improvement, have been well verified in plants (Csencsics et al., 2010; Lee et al., 2009; Takundwa et al., 2010). Despite their many advantages, SSR markers had not been developed or used in gerbera. Therefore, we developed and used SSRs to assess genetic diversity in gerbera accessions. Of these SSR markers, 75% were perfect type, 12.5% were imperfect and 12.5% were compound type. In the dendrogram shown in Figure 3, all 28 *G. hybrida* cultivars had no clear distinction between flower types. These differences could be attributed to the

low number of SSR markers employed.

In this study, we reported the successful development of eight polymorphic markers using 454-pyrosequencing technology. We also tested the applicability of these markers in *G. hybrida*. Although, only eight markers have been tested here, these microsatellite markers provide a useful tool for gerbera breeding programs. The com-prehensive dataset from this study will contribute to further genetic improvements, geno-mic information and gene discovery in gerbera. To our knowledge, this is the first report on micro-satellite markers for gerbera (*G. hybrida*) and it confirms the usefulness of this kind of marker as a tool for cultivar identification and the breeding   of
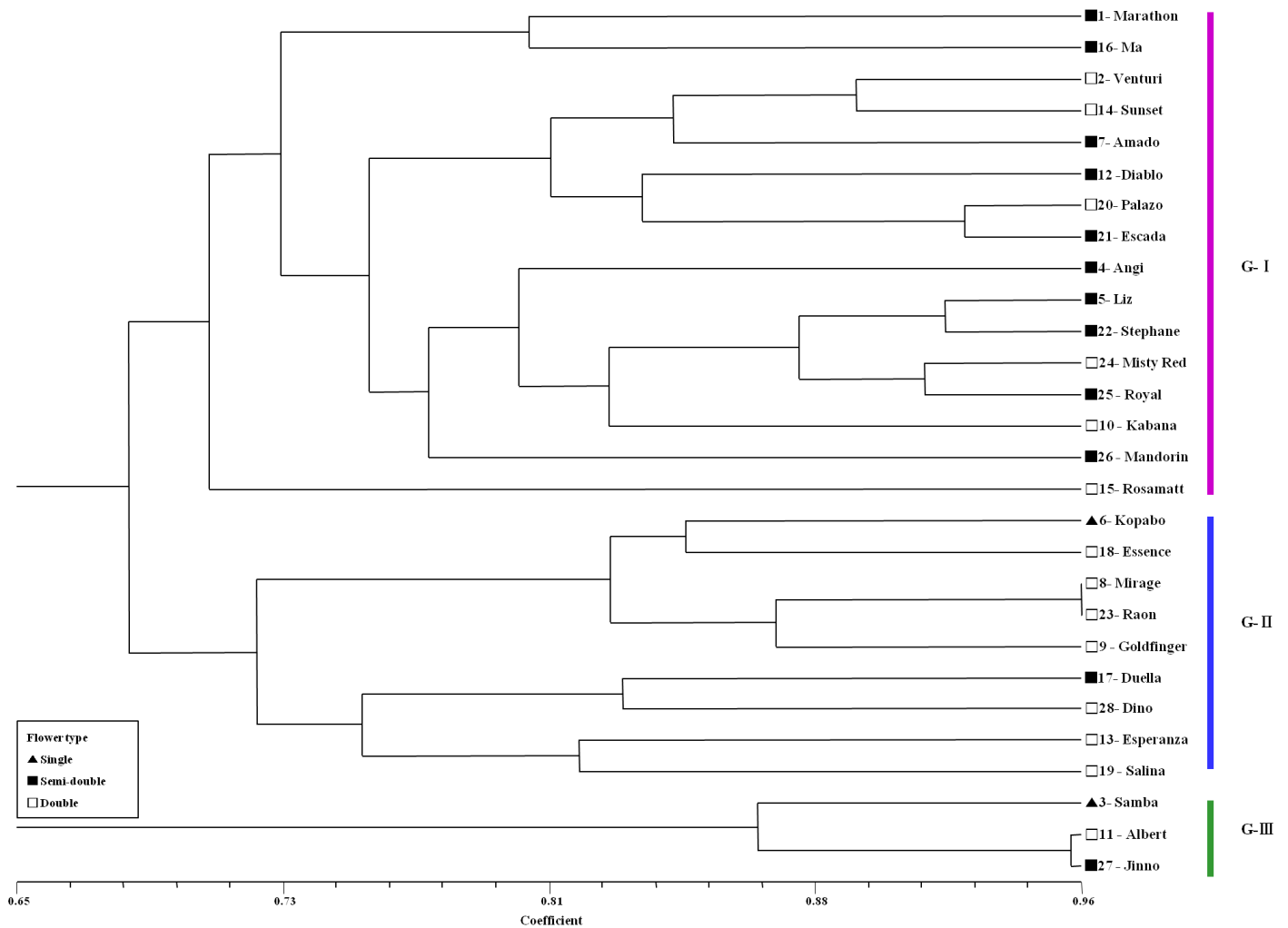
Figure 3. Dendrogram showing relationships among 28 *G. hybrida* accessions based on UPGMA of SSR profiles.

Table 5. Calculation of clonal diversity within groups.

Group	Number of samples	Number of genotypes	Nei's div[Z]
Group-I	16	14	0.98
Group-II	9	8	0.97
Group-III	3	3	1.00
Total	28	25	
Total diversity			0.99
Average diversity within populations			0.98

[Z]Nei's div: Nei's (1987) genetic diversity corrected for sample size.

new cultivars.

Agricultural Science, RDA, Republic of Korea.

REFERENCES

Algharim HJ, Almirall JR (2003). Development of microsatellite markers in Cannabis sativa For DNA typing and genetic relatedness analyses. Anal. Bioanal. Chem. 376: 1225-1233.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389-3402.

Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis* thaliana. Nature, 408: 796-815.

Ashkenazi V, Chani E, Lavi U, Levy D, Hillel J, Veilleux RE (2001). Development of microsatellite markers in potato and their use in phylogenetic and fingerprinting analyses. Genome, 44: 50-62.

Becker J, Heun M (1995). Barley microsatellites: allele variation and mapping. Plant Mol. Biol. 2: 835-845.

Benchimol LL, de Campos T, Carbonell SAM, Colombo CA, Chioratto AF, Formighieri EF, Gouvêa LRL, de Souza AP (2007). Structure of genetic diversity among common bean (*Phaseolus vulgaris* L.) varieties of Mesoamerican and Andean origins using new developed microsatellite markers. Genet. Resour. Crop Evol. 54: 1747-1762.

Bhatia R, Singha KP, Jhangb T, Sharmab TR (2009). Assessment of clonal fidelity of micropropagated gerbera plants by ISSR markers. Sci. Hortic. 119: 208-211.

Bonnal RJ, Severgnini M, Castaldi A, Bordoni R, Iacono M, Trimarco A, Torella A, Piluso G, Aurino S, Condorelli G, De Bellis G, Nigro V (2010). Reliable resequencing of the human dystrophin locus by universal long PCR and massive pyrosequencing. Anal. Biochem. 406: 176-184.

Chen C, Zhou P, Choi YA, Huang S. Gmitter FG (2006). Mining and characterizing microsatellites from citrus ESTs. Theor. Appl. Genet. 112: 1248-1257.

Choi SY, Kim TI, Shin HK, Kim HD, Joung HY (2001). Breeding of new gerbera 'Raon' of double type for cut-flower. J. Korean Soc. Hortic. Sci. 19(1): p. 130.

Conesa A, Götz S (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int. J. Plant Genomics, pp. 619-832.

Csencsics D, Brodbeck S, Holderegger R (2010). Cost-effective, species-specific microsatellite development for the endangered dwarf Bulrush (*Typha minima*) using next-generation sequencing technology. J. Hered. 101: 789-793.

Dellaporta SL, Wood J, Hicks JB (1983). A plant DNA minipreparation: version II. Plant Mol. Biol. Rpt. 4: 19-21.

Dice LR (1945). Measures of the amount of ecological association between species. Ecology, 26: 297-302.

Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2001). Assessment of genotypic variation among cultivated durum wheat based on EST-SSRs and genomic SSRs. Euphytica, 119: 39-43.

Gong Li, Deng Z (2010). EST-SSR markers for gerbera (*Gerbera hybrida*). Mol. Breed. 26: 125-132.

Kanwar JK, Kumar S (2008). *In vitro* propagation of Gerbera-a review. Hort. Sci. 35: 35-44.

Kim KW (2004). Developing one step program (SSR Manager) for rapid identification of clones with SSRs and primer designing. Master Thesis Seoul National University, the Republic of Korea.

Kumpatla SP, Mukhopadhyay S (2005). Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. Genome, 48: 985-998.

Lagercrantz U, Ellegren H, Andersson L (1993). The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res. 21: 1111-1115.

Lee JR, Hong GY, Dixit A, Chung JW, Ma KH, Lee JH, Kang HK, Cho YH, Gwag JG, Park YJ (2008). Characterization of microsatellite loci developed for *Amaranthus hypochondriacus* and their cross-amplifications in wild species. Conserv. Genet. 9: 243-246.

Lee RM, Thimmapuram J, Thinglum KA, Gong G, Hernandez AG, Wright CL, Kim RW, Mikel MA, Tranel PJ (2009). Sampling the waterhemp (*Amaranthus tuberculatus*) genome using pyrosequencing technology. Weed Sci. 57: 465-469.

Mata TLD, Segeren MI, Fonseca AS, Colombo CA (2009). Genetic divergence among gerbera accessions evaluated by RAPD. Sci. Hort. 121: 92-96.

Malausa T, Gilles A, Meglecz E, Blanquart H, Duthoy S, Costedoat C, Dubut V, Pech N, Castagnone-Sereno P, Delye C, Feau N, Frey P, Gauthier P, Guillemaud T, Hazard L, Le Corre V, Lung-Escarmant B, Male P, Ferreira S, Martin J (2011). High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. Mol. Ecol. Res. 11: 638-644.

Meirmans PG, Van Tienderen PH (2004). Genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. Mol. Ecol. Notes, 4: 792-794.

Messing J (1983). New M13 vectors for cloning. Method Enzymol. 101: 20-78.

Pettersson E, Lundeberg J, Ahmadian A (2009). Generations of sequencing technologies. Genomics, 93: 105-111.

Ruiz C, Breto MP, Ası′ns MJ (2000). An efficient methodology to identify sexual seedlings in citrus breeding programs using SSR markers. Euphytica, 112: 94-98.

Sato S, Hirakawa H, Isobe S, Fukai E, Watanabe A, Kato M, Kawashima K, Minami C, Muraki A, Nakazaki N, Takahashi C, Nakayama S, Kishida Y, Kohara M, Yamada M, Tsuruoka H, Sasamoto S, Tabata S, Aizu T, Toyoda A, Shin-I T, Minakuchi Y, Kohara Y, Fujiyama A, Tsuchimoto S, Kajiyama S, Makigano E, Ohmido N, Shibagaki N, Cartagena JA, Wada N, Kohinata T, Atefeh A, Yuasa S, Matsunaga S, Fukui K (2010). Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. DNA Res. 18(1): 65-76.

Schuelke M (2000). An economic method for the fluorescent labeling of PCR fragments. Nat. Biotechnol. 18: 233-234.

Sexton GJ, Frere CH, Dieters MJ, Godwin ID, Prentis PJ (2010). Development and characterization of microsatellite loci for *Khaya senegalensis* (Meliaceae). Am. J. Bot. 97: 111-113.

Takundwa M, Nepolo E, Chimwamurombe PM, Cullis AC, Kandawa-Schulz MA, Kunert K (2010). Development and use of microsatellites markers for genetic variation analysis, in the Namibian germplasm, both within and between populations of marama bean (*Tylosema esculentum*). J. Plant Breed. Crop Sci. 2: 233-242.

Tangphatsornruang S, Somta P, Uthaipaisanwong P, Chanprasert J, Sangsrakru D, Seehalak W, Sommanas W, Tragoonrung S, Srinives P (2009). Characterization of microsatellites and gene contents from genome shotgun sequences of mungbean [*Vigna radiata* (L.) Wilczek]. BMC Plant Biol. 9: p. 137.

Uzunova MI, Ecke W (1999). Abundance, polymorphism and genetic mapping of microsatellites in oilseed rape (*Brassica napus* L.). Plant Breed. 118: 323-326.

Wicker T, Schlagenhauf E, Graner A, Close T, Keller B, Stein N (2006). 454 sequencing put to the test using the complex genome of barley. BMC Genomics, 7: p. 275.

Yang P, Li X, Shipp MJ, Shockey JM, Cahoon EB (2010). Mining the bitter melon (*Momordica charantia* L.) seed transcriptome by 454 analysis of non-normalized and normalized cDNA populations for conjugated fatty acid metabolism-related genes. BMC Plant Biol. 10: p. 250.

Yu DY, Kotilainen M, Pollanen E, Mehto M, Elomaa P, Helariutta Y, Albert VA, Teeri TH (1999). Organ identity genes and modified patterns of flower development in *Gerbera hybrida* (Asteraceae). Plant J. 17: 51-62.

Zhu W, Ouyang S, Iovene M, O'Brien K, Vuong J, Jiang H, Buell CR (2008). Analysis of 90 Mb of the potato genome reveals conservation of gene structures and order with tomato but divergence in repetitive sequence composition. BMC Genomics, 9: p. 286.