Plant Omics Journal POJ
POJ 5(2):159-166 (2012) ISSN:1836-3644

Identification, characterization and validation of SSR markers from the gerbera EST database

Daiane de Pinho Benemann[1*], Luciana Nogueira Machado[1], Luis William Pacheco Arge[1], Valmor João Bianchi[1], Antonio Costa de Oliveira[2], Luciano Carlos da Maia[2], José Antonio Peters[1]

[1]Laboratory of Tissue Culture Plants. Institute of Biology, Department of Botany, Federal University of Pelotas, Pelotas, RS, Brazil. Campus UFPel S/N. Capão do Leão, CEP: 96010-900
[2]Plant Genomics and Breeding Center, FAEM, Federal University of Pelotas, Pelotas, RS, Brazil. Campus UFPel S/N. Capão do Leão, RS. CEP: 96010-900

***Corresponding author: daiane_bio@yahoo.com.br**

Abstract

Genic microsatellites, are currently among the best DNA markers because they often represent transcribed genes with putative function, and their cost is relatively low. In this study, 7,529 non-redundant (nr) ESTs from gerbera, obtained from public databases, were identified and characterized to validate their use as genic microsatellite markers. Of the assembled nr ESTs, 1,244 contained SSRs, and trinucleotides (GAA) were found to be the most abundant repeats at 42.3%, followed by dinucleotides (GA) at 25.3% and tetranucleotides (TTTG) at 6.2%. Gene ontology (GO) characterization was used to perform functional annotation of unigenes containing SSRs, of which 36.87% revealed significant sequence similarities with other species, indicating the top species with the highest similarity. A total of 647 primer pairs may be designed from the nr SSR-positive ESTs (excluding those with only mononucleotide repeats) for amplification of potential genic markers. A sample of 50 primer pairs was tested, and 17 were validated. The identified markers were highly polymorphic and able to differentiate 34 genotypes, representing the genetic variation within commercial and wild varieties. This study provides insight as to the frequency and distribution of SSRs in the gerbera transcriptome and demonstrates the successful development of genic SSRs. We expect that the potential markers described here can greatly expand the repertoire of DNA markers available for both breeding and genetic studies in *Gerbera spp*.

Keywords: Molecular markers, EST-SSR, *Gerbera hybrid.*
Abbreviations: SSR- Simple Sequence Repeats, EST- Expressed Sequence Tag, GO- Gene Ontology, PIC- Polymorphism Information Content, RAPD- Random Amplification of Polymorphic DNA, ISSR- Inter-Simple Sequence Repeat, He- Heterozygosity Expected, Ho- Heterozygosity Observed, UPGMA- Unweighted Pair Group Method with Arithmetic Mean, NCBI- National Center for Biotechnology Information, MAS- Marker Assisted Selection, UTR- Untranslated Region, CDS- Coding Sequence, nr- Non-Redundant.

Introduction

Simple Sequence Repeats (SSRs), also known as microsatellites, are DNA sequences that are formed by the arrangement of nucleotides through the combination of one to six base pairs that are repeated in tandem. These markers have been used in the analysis of genetic variation because they are consistent, robust and abundant in most eukaryote genomes (Morgante and Olivieri, 1993; Gupta et al., 2000; Tóth et al., 2000). As a result, microsatellites have become the most favored genetic markers used for applications such as plant breeding, assessing genetic diversity, constructing framework genetic maps, mapping useful genes, marker-assisted selection and comparative mapping studies (Gonzalo et al., 2005; Barakat et al., 2011). *Gerbera hybrida* belongs to the large sunflower family of plants, has gained increasing interest as a new model plant for developmental studies in Asteraceae and is one of the most important cut and pot flowers worldwide, ranking fifth, only after rose, carnation, chrysanthemum, and tulip, in the global cut flower trade (Bhatiaa et al., 2009; Buzgo et al., 2004; Teeri et al., 2006).
Gerbera is valued for its unique and attractive flower forms and bright colors. Pioneered in England in the late 19th century, gerbera breeding has been very popular in the Netherlands, Denmark, Germany, the United States, Israel and Japan (Rogers and Tjia 1990; Kloos et al., 2005). Nevertheless, expressed sequence tags (ESTs) for gerbera became available only recently in GenBank (Laitinen et al., 2005). For many plants, EST databases have become a valuable resource for *in silico* identification of SSRs, providing a cost-effective, labor-efficient approach for SSR marker determination (Varshney et al., 2005). Despite the apparent advantages for gerbera, the development of SSR marker databases has been slow for this important ornamental crop due to the limited number of microsatellites that have been reported to date (Gong and Deng, 2010).The protocols for recognizing SSR loci in a new species have historically been very labor-intensive. Nowadays, the use of bioinformatics tools helps to maximize the identification of biological data originating from whole genome sequence initiatives, consequently, the efficiency in the number of generated markers (Maia et al., 2008). The large amounts of sequence data generated from high-throughput genomic projects have required computational efforts to systematically mine them for microsatellites (Cardle et al., 2000; Temnykh et al., 2001; Kantety et al., 2002; Webster et al., 2002). The identification of EST-derived microsatellites, are more important than other based on assessing anonymous regions because they are functional markers (Varshney et al., 2002, 2005).To develop genic SSR markers for gerbera, 7,529 publically available non-redundant (nr) ESTs (http://www.ncbi.nlm.nih.gov) were explored with the following objectives: (1) to analyze the frequency and distribution of SSRs in gerbera transcripts, (2) to develop

novel EST–SSR primers, (3) to predict functional transcripts using Gene Ontology and (4) to validate EST–SSR primers via PCR experiments.

Results

Frequency and distribution of gerbera EST-SSRs

ESTs are essential for the discovery of genes and the detection of sequence features. However, they often represent partial and redundant short cDNA sequences, making their analysis difficult. To construct longer and less redundant sequence sets, the publicly available ESTs from *Gerbera hybrida* were assembled using CAP3 software (Huang and Madan, 1999). The initial number of EST sequences was 15.851, consisting of 7.254.970 bp. After running CAP3, that number was reduced to 7.529 contigs, composed of 3.941.688 bp. These 7.529 nr EST sequences were analyzed with the SSRLocator software (Maia et al., 2008). The resulting SSR loci were classified into two categories: SSR loci longer than 20 repeated units (Class I) and SSR loci equal to or longer than 12 but shorter than 20 repeated units (Class II) (Temnykh et al., 2001). A total of 16.6% of the identified EST-SSRs were grouped into Class I, while 83.4% were Class II. In general, trinucleotide repeats were the most abundant, representing 42.3% of the EST-SSRs, followed by dinucleotide repeats (25.3%) and then tetra-, hexa- and pentanucleotide repeats (16.6, 10.4 and 5.5%, respectively), as shown in Table 1. Among dinucleotide repeats, the GA motif (27.0%) was the most abundant, followed by CT (18.0%) and TC (13.0%), whereas GC (1.0%) was the least common (Fig 1A). As for the trinucleotide repeats, the GAA motif (6.5%) was the most abundant, followed by TAT (4.9%), TTC (4.8%) and TCT (4.0%), whereas the CTT motif (2.9%) was the least common (Fig 1B). Finally, regarding tetranucleotides, the motifs TTTG (6.2%), AAAC and TTTA (4.8%) were the most abundant (Fig 1C). Although the pentamer and hexamer repeats contained many different motifs, we found no significant patterns in their distribution.

Annotation and functional classification

The GO annotation results were grouped according to the following domains: molecular function, cellular component, and biological process (Fig 2). GO were identified for 569 transcripts according to cellular component (level 5), 995 transcripts for biological process (level 2) and 641 transcripts for molecular function (level 5). In the case of the cellular component domain, 11 categories were assigned with the majority of transcripts belonging to either the “intracellular part” (59.2%) or “plasma membrane” (14.4%) (Fig 2A). For the biological process ontology, the EST-SSRs were sorted into 14 categories with the majority of sequences representing genes involved in “cellular processes” (28.8%) and “metabolic processes” (25.5%) (Fig 2B). In the molecular function ontology, EST-SSRs fell into 12 categories with the most sequences assigned to “binding” (46.4%) and “catalytic activity” (34.1%) (Fig 2C). Finally, GO analysis indicated that the best results from BLASTX searches were most closely related to *Vitis vinifera*, followed by *Arabidopsis thaliana*, *Populus trichocarpa* and *Oryza sativa*.

EST-SSR primer validation and genetic analysis

From the sequences analyzed, 50 primer sets were selected and designed by the SSRLocator software. Forty-two sets were designed for the amplification of Class I EST-SSRs, and eight sets were designed for Class II EST-SSRs. These primers were further tested for 34 gerbera genotypes (Table 2). Polymorphisms were identified for 20 out of the 50 loci analyzed (40.0%), and 3 loci resulted in weak amplifications or nonspecific amplicons. Generally, inconsistent amplification or amplification failure from EST-SSR loci may be the result of factors such as introns that are too large to permit efficient amplification, poor quality sequences for primer design, or mutational substitutions (insertions or deletions) within the priming sites (De Jong et al., 2003). Marker evaluation details are provided in supplementary data. Amplification success rates were maximized in cases where the primer sets were anchored in trimer sequences (52.94%) followed by dimer (41.17%) and tetramer (5.88%) anchors. The number of alleles detected per locus in the present case ranged from five to seven, with an average of 5.94 alleles detected per locus. The He ranged from 0.68 to 0.77 (with an average of 0.72) and the Ho ranged from 0.4 to 0.64 (with an average of 0.52) (Table 3). The PIC ranged from 0.61 to 0.73, with an average of 0.67 and a cophenetic coefficient value of 0.87. The genetic distance (D) among the different genotypes using the 17 EST-SSRs was reproduced in a UPGMA dendrogram (Fig 3) according to the original data obtained in the similarity matrix and based on the genetic distances among the genotypes. The dendrogram illustrated three significant main groups, using the average similarity as a cutoff value. Group 1 included four genotypes: three from the cultivar Igor and one from the cultivar Cariba-P22. These cultivars both exhibit semi-folded inflorescence; however, Igor has pink flowers, while Cariba has red. Group 2 consisted only of Terra fame cultivars, which have yellow and simple-type inflorescence. Group 3 was composed of Golden Kozak, Cosmo, Monique, G32, Deranagem and Orça-P40. The remaining cultivars, including Orça-P42, Pacific, Mystique, Pink Elegance and wild-type, did not form clusters with other genotypes, showing similarity levels below the average.

Discussion

SSR frequency and distribution

Theoretically, the frequency of a repeat should decrease as its length increases. However, we found trimers to be the most commonly occurring SSRs, followed by dimers, tetramers, hexamers and pentamers (Table 1). This finding is consistent with the results reported for other plant species (Eujayl et al., 2004; La Rota et al., 2005; Luro et al., 2008; Varshney et al., 2002; 2005; Zeng et al., 2010). Metzgar et al. (2000) explained this result by suggesting that the abundance of expressed trimers is a consequence of negative selection against mutations that alter the reading frame in coding regions, thus favoring the presence of repeats in triplet. Accordingly, dinucleotide repeats are reported to be more polymorphic among SSRs in plant species (Lagercrantz et al., 1993; Morgante and Olivieri 1993; Li et al., 2002), while trinucleotide repeats are over-represented in coding sequences but occur less frequently than mono- and dinucleotide repeats in non-coding regions (Tòth et al., 2000; Gao et al., 2003). Furthermore, microsatellite distribution seems to differ between intergenic and intronic sequences (Morgante et al., 2002). As shown in Fig 1, the most dominant dinucleotide repeat motifs in the gerbera nr EST were GA and CT (27.0% and 13.0%, respectively). These motifs were also the most frequently observed SSRs in other plants (Gao et al., 2003; Saha et al., 2004; Scott et al., 2000; Thiel et al., 2003), and may correspond to the GAU, GAC,

Table 1. Frequencies of Class I and Class II repeats in gerbera.

Type		Di-	Tri-	Tretra-	Penta-	Hexa-	Total
Class I	Occurrence	134	25	9	11	27	206
	%	10.8	2	0.7	0.9	2.2	17
Class II	Occurrence	181	501	197	57	102	1.038
	%	14.5	40.3	15.8	4.6	8.2	83
Total	Occurrence	315	526	206	68	129	1.244
	%	25.3	42.3	16.6	5.5	10.4	-

Table 2. Gerbera access analyzed through EST-SSR markers, their main flower type and inflorescence color.

Number access	Access name	Flower type	Inflorescence color
P1	Igor	Semi folded	Pink, light center
P2	Terra fame	Simple	Yellow, light center
P4	Golden G.	Semi folded	Yellow, black center
P5	Igor	Siemi folded	Pink, light center
P6	Kozak	Semi folded	Dark orange, black center
P7	Wild type	Simple	Red
P8	Pink elegance	Semi folded	Pink, light center
P9	Golden G.	Semi folded	Yellow, black center
P10	Golden G.	Semi folded	Yellow, black center
P12	Deranagem	Semi folded	Pink, black center
P13	Terra fame	Simple	Yellow, light center
P14	Golden G.	Semi folded	Yellow, black center
P15	Deranagem	Semi folded	Pink, black center
P16	Cariba	Semi folded	Red, light center
P17	Igor	Semi folded	Pink, light center
P18	Terra fame	Simple	Yellow, light center
P19	Mystique	Simple	Orange, black center
P20	Golden G.	Semi folded	Yellow, black center
P21	Willd type	Simple	Orange
P22	Cariba	Semi folded	Red, light center
P23	Terra fame	Simple	Yellow, light center
P24	Grolden G.	Semi folded	Yellow, black center
P25	Terra fame	Simple	Yellow, light center
P26	Golden G.	Semi folded	Yellow, black center
P27	Terra fame	Simple	Yellow, light center
P28	Cosmo	Simple	Yellow, black center
P29	Terra fame	Simple	Yellow, light center
P31	Terra fame	Simple	Yellow, light center
P32	G32	Simple	Yellow
P34	Monique	Semi folded	Red, light center
P40	Orça	Simple	Completely white
P42	Orça	Simple	Completely white
P43	Pacifc	Simple	Completely white
P45	Pacifc	Simple	Completely white

GAA and GAG codons, which encode the amino acids aspartate and glutamate. A high abundance of $(GA)_n$ microsatellites compared to other dinucleotide SSRs has previously been observed in plant genomes, including *Oryza sativa*, *Aegilops triuncialis, Arabidopsis thaliana, Brassica napus, Glycine max* and *Helianthus annuus* (Akkaya et al., 1992; Guyomarc et al., 2002; Gupta and Varshney, 2000; Suwabe et al., 2002; Uzunova and Ecke, 1999). Earlier studies on plant $(GA)_n$ microsatellites also demonstrated that they are well distributed throughout the genome, thus ensuring good genome coverage. CT repeats have been widely targeted for EST-SSR marker development in plants because, in addition to being highly abundant, they are often highly polymorphic, more abundant in UTRs than CDSs, seldom associated with transposons, are consistently amplified and yield robust SSR markers (Morgante et al., 2002; Temnykh et al., 2001). The most abundant tri-nucleotide repeat motif detected in the present study was GAA/TAT/TCT (6.5%, 4.9% and 4.0%, respectively), coding for glutamic acid, tyrosine and phenylalanine, respectively. These motifs have also been observed in *Arabidopsis*, almond, peach and rose (Jung et al., 2005). GAA and TTC motifs are the most abundant repeats in *Camellia sinensis* (Sharma et al., 2009), and the TTC motif occurs frequently in *Coffea canephora* (Ponset et al., 2006).

Gerbera EST annotation

Functional annotation allows the categorization of genes into functional classes, which can be very useful for understanding the physiological meaning of large numbers of genes and assessing the functional differences between subgroups of sequences (Nelson and Shoemaker, 2006). Analysis of the distribution of specific terms was performed using the Blast2GO software (Conesa et al., 2005).

The majority of unigenes identified in this study shared significant similarity with genes from other plant species. The most similarities were found with *Vitis vinifera*,

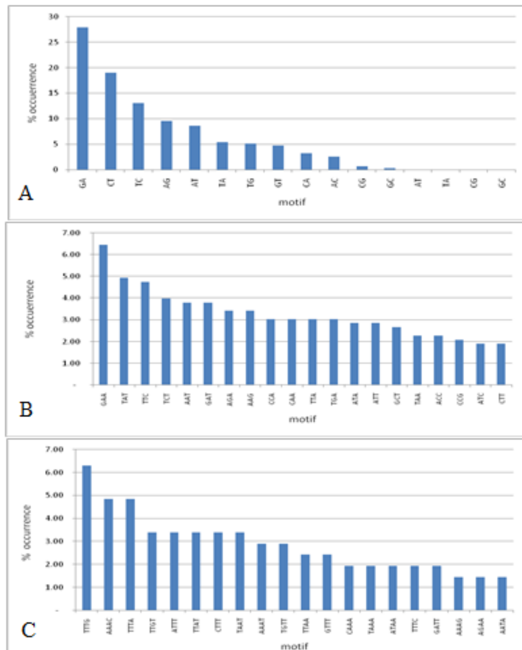

Fig 1. Frequency of the SSR motifs in EST sequences from gerbera. (A) Dimers, (B) trimers and (C) tetramers.

Arabidopsis thaliana, and *Populus trichocarpa*. However, these species are all core eudicots and rosids, whereas gerbera belongs to the core eudicot clade asterides (Bremer et al., 2003). Moccia et al. (2009) reached the same conclusion and proposed that the similarity can be attributed to the fact that these species have fully sequenced genomes and large EST databases. The suggestion that these similarities do not reflect phylogenetic proximity between species was further verified in the current study. It may be interesting to test the primers designed here in the related species to confirm or rule out these proposed similarities.

***Genetic* variability**

Historically, molecular markers such as RAPD, ISSR and EST-SSR have all been used to estimate the genetic diversity in different cultivars of gerbera (Bhatia et al., 2009; Gong and Deng, 2010; Mata el al., 2009; Rezende et al., 2009). Assessment of genetic diversity through molecular markers is useful for the conservation of genetic resources, cultivar identification, and parental selection for hybridization in breeding programs (Garner et al., 2004). Therefore, the selection of a particular type of molecular marker is important and depends on the intended use (Gupta et al., 1996). SSRs are among the best DNA markers because they are highly polymorphic, highly abundant, genetically co-dominant, and analytically simple (Gupta and Varshney, 2000; Varshney et al., 2005). In recent years, EST-SSRs are being used more frequently for genotyping of natural or breeding populations due to their inexpensive developmental costs (Kota et al., 2003; Varshney et al., 2005). Moreover, because EST-SSRs are derived from transcribed sequences, EST-SSR markers generally demonstrate higher levels of transferability across species than do genomic SSR

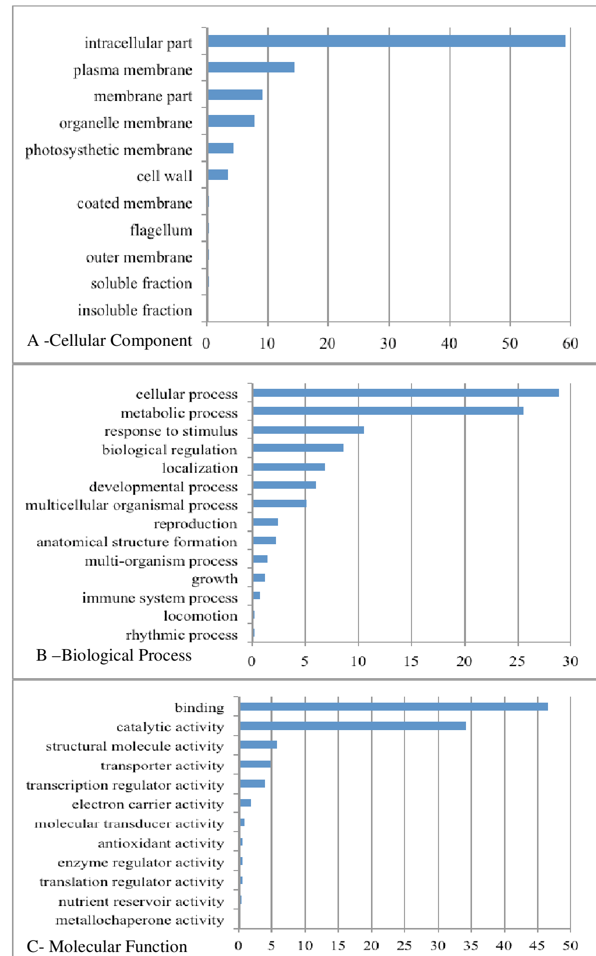

Fig 2. Gene Ontology (GO) classification of EST-SSR sequences from gerbera unigenes. (A) Cellular Component, (B) Biological Process and (C) Molecular Function.

markers. EST-SSRs also have the potential for direct linkage association to functional genes encoding important horticultural traits. The set of 17 EST-SSR primers used in this study revealed a total of 101 polymorphic alleles, with an average of 5.94 per locus. Similarity studies in gerbera by Chung et al. (2001) and Rezende et al. (2009) detected an average of 1.6 (analyzing 24 accesses) and 2.95 (analyzing 6 accesses) polymorphic bands per primer set, respectively. These studies used RAPD markers and found fewer alleles than the current study. Mata et al. (2009) identified an average of 6.1 polymorphic bands by analyzing 42 accesses of gerbera, also using RAPD markers. Finally, an average of 2.7 polymorphic alleles was found in 7 accesses of gerbera using EST-SSR as the molecular markers (Gong and Deng, 2010). The differences in identified polymorphisms could be attributed to the different marker techniques in some cases and to the different population sizes in others (Gong and Deng, 2010). Therefore, the present study is the most complete EST-SSR marker survey in this species to date. Usually, the average number of alleles per locus increases as the number of samples or SSRs examined increases. An important characteristic of a genetic marker is its heterozygosity; that is, the probability that a random individual is heterozygous for a given marker. In the

Table 3. Diversity statistics for the 17 EST-SSR loci in gerbera, described by the number of alleles per locus, expected (He) and observed (Ho) heterozygosities, endogamy coefficient (F), and polymorphism information content (PIC) observed in the 34 gerbera genotypes.

Locus name	N° alleles	He	Ho	F	PIC
GERB1	7	0.73	0.55	0.23	0.68
GERB4	6	0.71	0.4	0.34	0.67
GERB9	6	0.77	0.44	0.42	0.73
GERB10	7	0.73	0.55	0.23	0.68
GERB12	7	0.76	0.55	0.26	0.72
GERB14	6	0.72	0.51	0.28	0.67
GERB17	6	0.73	0.5	0.32	0.69
GERB21	6	0.72	0.52	0.27	0.68
GERB22	5	0.67	0.64	0.03	0.61
GERB25	5	0.69	0.57	0.17	0.64
GERB30	6	0.72	0.51	0.29	0.68
GERB34	6	0.72	0.58	0.19	0.68
GERB35	5	0.68	0.43	0.36	0.62
GERB40	7	0.75	0.48	0.35	0.71
GERB41	6	0.7	0.51	0.27	0.65
GERB42	5	0.68	0.52	0.23	0.63
GERB44	5	0.68	0.51	0.24	0.62
Total	101				
Mean	5.94	0.72	0.52	0.26	0.67

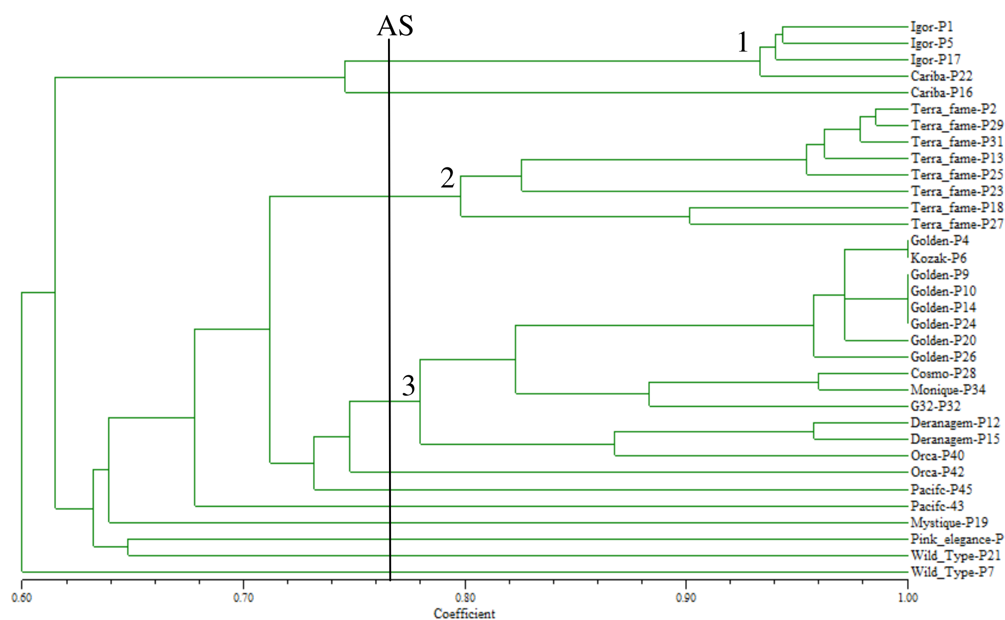

Fig 3. Dendrogram of the 34 gerbera genotypes obtained from UPMGA cluster analysis, using the Simple Matching coefficient after amplification with 17 different EST-SSR primer pairs. The dendrogram illustrated 3 significant main groups, using the average similarity as a cutoff value.

present study, the He and Ho were 0.72 and 0.52, respectively. Heterozygosity in *Prunus* was found to be 0.66 and 0.46 for He and Ho, respectively (Bouhadida et al., 2009). Other values obtained ranged from 0.35 to 0.46 for He in *Epimedium sagittatum* and *Eriobotrya japonica*, respectively. The same studies detected values ranging from 0.65 to 0.51 for Ho in the same species (Soriano et al., 2005; Zeng et al., 2010). High allele number and high heterozygosity obtained in the present study reflect the ability of SSR markers to provide unique genetic profiles for individual plant genotypes. Genetic analysis of the accesses did not detect differences between the genotypes 'Golden (P4)' and 'Kozak (P6)', which are both semi-folded with black centers. Similarly, other Golden genotypes (P9, P10, P14 and P24) showed completely identical patterns. This analysis is shown in Fig 3, where most commercially cultivated genotypes clustered together, but the wild-type genotype was the most dissimilar. Analysis of a sub-set of the genotypes used in this study was also performed by Mata et al. (2009) using RAPDs. In that study, lower similarities were detected

between Golden and Igor (0.69) and Cosmo and Orça (0.69), and the genetic similarity between both groups was 0.60. Kozak had a genetic similarity of 0.58 between both groups. In our results, Golden and Igor did not cluster together. Although the similarity between Golden and Igor was 0.62, the similarity between Cosmo and Orça was 0.73. Kozak exhibited a similarity of 1.00 with Golden; in addition, Kozak or Golden showed a similarity of 0.82 with Cosmo, 0.78 with Orça and 0.62 with Igor. These small differences could be due to the use of different molecular markers, the number of accesses analyzed and the number of primers used. Remarkably, all the accesses for Igor clustered together; in addition, the accesses for Terra fame, Golden, Pacific, Orça and Deranagem exhibited similar findings. These results show the efficiency of primers designed by the program SSRLocator.

In summary, the present study attempts to ascertain the frequency and distribution of SSRs in the gerbera EST database and develop those EST-SSRs for use in genetic studies. The resulting EST-SSR set is a valuable tool that has been made publicly available for further genetic and genomic applications. Seventeen markers were developed and experimentally validated for genetic analysis in different cultivars of *Gerbera* spp. The markers were highly polymorphic and were able to differentiate the gerbera accesses. The use of these markers would reduce the cost and therefore facilitate genetic distance assessments, gene mapping and possible marker-assisted selection (MAS) in *Gerbera* ssp. The functional categorization of these markers corresponded to many genes with biological, cellular and molecular functions, thus providing an opportunity to investigate the consequences of SSR polymorphisms on gene function and the possible impact on phenotypic differences. Application of this knowledge may help to directly drive crosses to obtain new gerbera hybrids.

Materials and methods

Plant material

For the present study, 34 gerbera genotypes were kindly supplied by the gerbera germplasm bank of Pró-clone (São Paulo, Brazil) and are described in Table 2.

EST-SSR characterization and primer design

All ESTs related to *Gerbera hybrida* available from the NCBI (http://www.ncbi.nlm.nih.gov) were downloaded in FASTA format. Redundancy was eliminated using CAP3 software (Huang and Madan, 1999). SSRLocator software (Maia et al., 2008) was used to find SSR loci and the options (types/motifs) were adjusted to locate dimers, trimers, tetramers, pentamers, and hexamers with respective minimums of 6, 4, 3, 3 and 3 repeats. Parameters for primer design were defined as follows: the primer length was between 18-22 with an optimum of 20 bp, the PCR products were from 100 to 300 bp, the primer TM was between 58°-64°C with an optimum of 60°C and the GC content ranged from 45-70%. The primers were synthesized by Invitrogen Biotechnology Co.,Porto Alegre- RS-Brazil.

Functional characterization

Functional annotation of gerbera markers was obtained from GenBank using the BLASTX algorithm against the nr database (Altschul et al., 1997) and further classified by gene ontology (Ashburner et al., 2000). Gene Ontology (GO) terms were extracted from the best homologous hit. The Blast2GO (B2G) term browser was used to define molecular functions, which describe activities, such as catalytic or binding at the molecular level, cellular component, which predicts the locations of gene products at the level of subcellular structures and macromolecular complexes, and biological processes, which describe the biological goals accomplished by one or more ordered assemblies of molecular functions. The program extracts GO terms from each obtained hit by mapping to existing annotation associations.

DNA extraction and PCR experiments for EST-SSR loci analysis

Young leaves were collected from 30 day-old plants. Genomic DNA was extracted from frozen leaves as described by Doyle and Doyle (1987). DNA concentration and quality were evaluated by electrophoresis in 0.8% agarose gel. For SSR markers, PCR amplification was performed in a 25 µL reaction containing 10 ng of genomic DNA, 1.0 U of Taq polymerase (Invitrogen), 0.5 µM of each primer (described in the supplementary material), 2.5 mM of each dNTP (Invitrogen), 1.5 mM of $MgCl_2$ and 2.5 µL of 10x buffer reaction. The PCR reactions were performed in an ICycler (Bio-Rad Laboratories, Hercules, CA, USA) as follows: 4 min at 94 °C followed by 35 cycles of 45 s at 94 °C, 45 s at the specific annealing temperature, and 90 s at 72 °C, with a final extension of 5 min at 72 °C. PCR reaction products were denatured by the addition of ½ volume of 95% formamide/dye solution and then heated for 5 min at 94 °C. After being chilled on ice, 4.5 µl of the denatured preparation was loaded on a 6% denaturing polyacrylamide gel containing 7.5 M urea and in 1X TBE running buffer. Following electrophoresis, the gels were silver-stained following the protocol described in Bassam et al. (1991). Fragment sizes were estimated with 25 bp and 100 bp DNA Ladder sizing markers (Invitrogen).

Genetic **variability** *and data analysis*

Polymorphisms were evaluated based on their presence (1) or absence (0), and the data were entered in a binary data matrix as discrete variables. Null alleles were assigned to genotypes that were confirmed to have no amplification products under standard conditions. A Simple Matching coefficient (Sokal and Michener, 1958) was used to calculate genetic similarity, and a dendrogram was developed using UPGMA. The computer package NTSYS-pc Ver. 2.1 (Rohlf et al., 2000) was used for cluster analysis and matrix correlation. The robustness of the clustering tree was evaluated by bootstrapping (1.000 replicates). Expected heterozygosity (He) was calculated as $H_e = 1 - \sum p_i^2$, where p_i is the frequency of the *i*th allele. Observed heterozygosity (Ho) was calculated as the number of heterozygous genotypes divided by the total number of genotypes. The fixation index was calculated as $F = 1 - H_o / H_e$ (Nei 1973; Wright, 1965; Botstein, 1980). Both bootstrap and heterozygosity analyses were performed using the Genes software package (Cruz et al. 1998). Polymorphic information content (PIC) was calculated as $PIC = 1 - \sum p_i^2$, where pi is the frequency of the ith allele of each SRR marker (Anderson et al., 1993).

Acknowledgements

The first author would like to thank CAPES for the PhD scholarship, Professor Odir Dellagostin who kindly provided the primers and Monique Seregen for kindly supplying 34 gerbera genotypes from the gerbera germplasm bank of the Pró-clone business, São Paulo-Brazil.

References

Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. Genome 36:181-186

Akkaya, MS, Bhagwat AA, Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131–1139

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

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. (2000) Gene ontology: Tool for the unification of biology. Nat Genet 25:25-29

Barakat MN, Al-Doss AA, Elshafei AA, Moustafa KA 2011 Identification of new microsatellite marker linked to the grain filling rate as indicator for heat tolerance genes in F2 wheat population. Aust J Crop Sci 5:104-110

Bassam BJ, Caetanoanolles G, Gresshoff PM (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal Biochem 196:80-83

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

Bremer B, Bremer K, Chase MW, Reveal JL, Soltis DE, Soltis PS, Stevens PF, Anderberg AA, Fay MF, Goldblatt P, et al. (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Bot J Linn Soc 141:399-436

Botstein D, White RL, Skolnick M, and Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Bouhadida M, Casas AM, Moreno MA, Gonzalo MJ, Arús P, Moreno MA, Gogorcena Y (2009) Molecular characterization nad genetic diversity of *Prunus* rootstoks. Sci Hort 120: 237-245

Buzgo M, Soltis DE, Soltis PS, Ma H (2004) Towards a comprehensive integration of morphological and genetic studies of floral development. Trends Plant Sci 9: 164–173

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 56:847–854

Chung Y, Kim H, Kim K, Park S, Yi Y, Lee J, Kwon O (2001) Morphological characteristics and genetic variation of gerbera (*Gerbera hybrida* hort). Journal of Plant Biotechnology 3:145-149.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatic 21:3674-3676

Cruz, CD (1998) Programa Genes – Aplicativo computacional em estatística aplicada à genética. Genet Mol Biol 21:135–138

De Jong EV, Guthridge KM, Spangenberg GC, Forster JW (2003) Development and characterization of EST-derived simple sequence repeat (SSR) markers for pasture grass endophytes. Genome 46:277-290

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

Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC, Mian MAR: Medicago truncatula (2004) EST-SSRs reveal cross-species genetic markers for Medicago spp. Theor Appl Genet 108:414-422

Garner EC, Campbell CS, Mullins RD (2004) Dynamic instability in a DNA-segregating prokaryotic actin homolog. Science 306:1021-1025

Gao LF, Tang JF, Li HW, Jia JZ (2003) Analysis of microsatellites in major crops assessed by computational and experimental approaches. Mol Breed 12:245–261

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

Gonzalo MJ, Oliver M, Garcia MJ, Monfort A, Dolcet SR, Katzir N, Arus PM (2005) A Simple-sequence repeat markers used in merging linkage maps of melon (*Cucumis melo* L.). Theor Appl Genet 110:802-811

Gupta PK, Balyan IS, Sharma PC, Ramesh B (1996) Microsatellites in plants: a new class of molecular markers. Curr Sci 70:45–54

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163–185

Guyomarc H, Sourdille P, Charmet G, Edwards KJ and Bernard M (2002) Characterization of polymorphic microsatellite markers from *Aegilops tauschii* and transferability to the D-genome of bread wheat. Theor Appl Genet 104:1164-1172

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

Jung S, Abott A, Jesudurai C, Tomkins J, Main D (2005) Frequency, type, distribution and annotation of simple sequence repets in Rosaceae ESTs. Funct Integr Genomics 5:136-143

Kantety RV, Rota ML, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48:501–510

Kota R, Rudd S, Facius A, Kolesov G, Thiel T, Zhang H, Stein N, Mayer K, Garner A (2003) Snipping polymorphisms from large EST collections in barley (*Hordeum vulgare* L). Mol Genet Genomics 270:224–233

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

Laitinen RAE, Immanen J, Auvinen P, Rudd S, Alatalo E, Paulin L, Ainasoja M, Kotilainen M, Koskela S, Teeri TH, Elomaa P (2005) Analysis of the floral transcriptome uncovers new regulators of organ determination and gene families related to flower organ differentiation in *Gerbera hybrida* (Asteraceae). Genome Res 15:475–486

La Rota M, Kantety RV, Yu JK, Sorrells ME (2005) Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. BMC Genomics 6:23

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions, and mutational mechanisms (a review). Mol Ecol 11:2543–2565

Luro FL, Costantino G, Terol J, Argout X, Allario T, Wincker P, Talon M, Ollitrault P, Morillon R (2008) Transferability of the EST-SSRs developed on *Nules clementine* (*Citrus clementina* Hort ex Tan) to other Citrus species and their effectiveness for genetic mapping. BMC Genomics 9:287

Maia LC da, Palmieri DA, de Souza VQ, Kopp MM, de Carvalho FI and Costa de Oliveira A (2008) SSR Locator: Tool for simple sequence repeat discovery integrated with primer design and PCR simulation. Int J Plant Genomics ID 412696

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

Meztgar D, Bytof J, Wills W (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res. 10:72-80

Moccia MD, Oger-Desfeux C, Marais GAB, Widmer A (2009) A White Campion (*Silene latifolia*) floral expressed sequence tag (EST) library: annotation, EST-SSR characterization, transferability, and utility for comparative mapping. BMC Genomics. 10: 243

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet. 30:194-200

Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J 3:175–183

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Soc. USA. 70: 3321–3323

Nelson RT, Shoemaker R (2006) Identification and analysis of gene families from the duplicated genome of soybean using EST sequences. BMC Genomic 7:204

Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, Kochko A, Hamon P (2006) SSR mining in coffee tree EST satabases: potential use of EST-SSRs as markers for the Coffea genus. Mol Gen Genomics 276:436-449

Rezende RKS, Paiva LV, Paiva R, Junior AC, Torga PP, Masetto TE (2009) Genetic divergence among cultivars of gerbera using RAPD markers. Ciência Rural, Santa Maria. 39:2435-2440

Rogers MN, Tjia BO (1990) *Gerbera production for cut flowers and pot plants*. Portland, Oregon: Timber Press. 116p.

Kloos WE, George CG, Sorge LK (2005) Inheritance of powdery mildew resistance and leaf macrohair density in *Gerbera hybrida*. HortScience 40:1246–1251

Rohlf FJ (2000) NTSYS-pc Numerical Taxonomy and Multivariate analysis System. Verson 2.1 Exeter Software, Setauket, NY

Saha MC, Mian MA, Eujayl I, Zwonitzer JC, Wang L, May GD (2004) Tall fescue EST-SSR markers with transferability across several grass species. Theor Appl Genet 109:783–791

Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. Theor Appl Genet 100:723–726

Sharma RK, Bhardwaj P, Negi R, Mohapatra T, Ahuja PS (2009) Identification, characterization and utilization of unigene derived microsatellite markers in tea (*Camellia sinensis* L.). *BMC* Plant Biol 9:53

Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. The University of Kansas Science Bulletin **28**. 1409-1438

Soriano JM, Romero C, Vilanova S, Llácer G, Badenes ML (2005) Genetic diversity of loquat germplasm (*Eriobotrya japonica* (Thunb) Lindl) assessed by SSR markers. Genome 48:108–114

Suwabe K, Iketani H, Numone T, Kage T and Hirai M (2002) Isolation and characterization of microsatellites in *Bassica rapa* L. Theor Appl Genet 104:1092-1098

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (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

Teeri TH, Elomaa P, Kotilainen M, Albert VA (2006) Mining plant diversity: Gerbera as a model system for plant developmental and biosynthetic research. BioEssays 28:756–767

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

Tóth G, Gáspári Z, Jurka J (2000) Microsatellites in Different Eukaryotic Genomes: Survey and Analysis. Genome Res 10:967–981

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

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55

Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002) In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 7:537-546

Webster M, Smith N, Ellegren H (2002) Microsatellite evolution inferred from human–chimpanzee genomic sequence alignments. Proc Natl Acad Sci USA 99:8748–8753

Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution. 19: 395–420

Zeng S, Xiai G, Guo J, Fei Z, Xu Y, Roe B, Wang Y (2010) Development of EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* (Sieb. Et Zucc.) Maxim. BMC Genomics 11:94