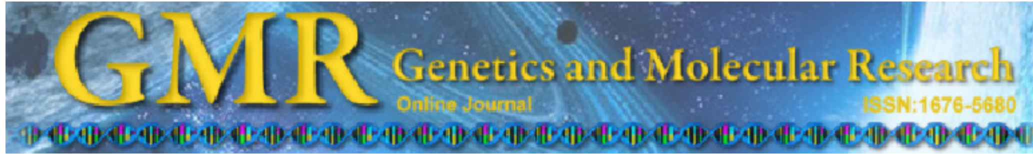

Characterization of EST-derived and non-EST simple sequence repeats in an F_1 hybrid population of *Vitis vinifera* L.

E. Kayesh[1,2], N. Bilkish[1], G.S. Liu[3], W. Chen[1], X.P. Leng[1] and J.G. Fang[1]

[1]College of Horticulture, Nanjing Agricultural University, Nanjing, China
[2]Department of Horticulture,
Bangabanhu Sheikh Mujibur Rahman Agricultural University,
Gazipur, Bangladesh
[3]College of Horticulture, Qing Dao Agricultural University, Qing Dao, China

Corresponding author: J.G. Fang
E-mail: fanggg@njau.edu.cn

Genet. Mol. Res. 13 (1): 2220-2230 (2014)
Received January 4, 2013
Accepted June 14, 2013
Published March 31, 2014
DOI http://dx.doi.org/10.4238/2014.March.31.2

ABSTRACT. Among different classes of molecular markers, expressed sequence tags (ESTs) are a new resource for developing simple sequence repeat (SSR) functional markers for genotyping and genetic mapping in F_1 hybrid populations of *Vitis vinifera* L. Recently, because of the availability of an enormous amount of data for ESTs in the public domain, the emphasis has shifted from genomic SSRs to EST-SSRs, which belong to transcribed regions of the genome and may have a role in gene expression or function. The objective of this study was to assess the polymorphisms among 94 F_1 hybrids from "Early Rose" and "Red Globe" using 25 EST-derived and 25 non-EST SSR markers. A total collection of 362,375 grape ESTs that were retrieved from the National Center for Biotechnology Information (NCBI) and 2522 EST-SSR sequences were identified. From them, 205 primer pairs were randomly selected, including 176 pairs that were EST-derived and 29 non-EST SSR primer pairs, for polymerase chain reaction amplification. A total

 ©FUNPEC-RP www.funpecrp.com.br

of 131 alleles were amplified using 50 pairs of primers; 78 alleles were amplified using EST-derived SSR primers and 53 were from non-EST SSR primers. At most, 6 and 5 alleles were amplified by EST-derived and non-EST SSR primers, respectively. The EST-derived SSR markers showed a maximum polymorphic information content (PIC) value of 1 and a minimum of 0.33 while non-EST SSR markers had maximum and minimum PIC values of 1 and 0.25, respectively. The average PIC value was 0.56 for EST-derived SSR markers and 0.45 for non-EST SSR markers.

Key words: Grapevine; non-EST SSR; Polymorphism;
Expressed sequence tag (EST)-derived simple sequence repeat (SSR)

INTRODUCTION

Among different classes of molecular markers, simple sequence repeats (SSRs) are the most suitable for studying polymorphisms because of their ease in handling, reproducibility, multiallelic nature, co-dominant inheritance, relative abundance, and genome-wide coverage (Powell et al., 1996). Recently, because of the availability of an enormous amount of data for expressed sequence tags (ESTs) in the public domain, the emphasis has shifted from genomic SSRs to EST-SSRs, which belong to transcribed regions of the genome and may have a role in gene expression or function.

EST projects have been initiated for numerous plant and animal species, generating large amounts of sequence information that can be used for gene discovery, functional genetic studies, and marker development (Pashley et al., 2006). ESTs were used for the first time in 1991 by Adams et al. as a means of gene discovery in the human brain. Since then, ESTs have played an important role in functional genomic research for the discovery of new functional genes other than whole-genome approaches (Chen et al., 2005; Yamada-Akiyama et al., 2009; Zhao et al., 2009).

The availability of ESTs greatly accelerates the systematic identification of SSRs and corresponding marker development based on computational approaches (Varshney et al., 2002; Gao et al., 2003; Thiel et al., 2003; Chen et al., 2006). EST-derived SSRs have been well documented in some plant species including *Arabidopsis* (Depeiges et al., 1995), sugarcane (Cordeiro et al., 2001), cereal species (Kantety et al., 2002), cacao (Lima et al., 2008), and rubber tree (Feng et al., 2009). Recently, many EST libraries of a wide range of plant species have been constructed for genes involved in plant growth and differentiation (Matsuoka et al., 2004), biochemical pathways (Remy and Michnick, 2004; Urbanczyk-Wochniak and Sumner, 2007), secondary metabolism (Park et al., 2004), and responses to environmental stresses and pathogen attack (Sugui and Deising, 2002). By July 1, 2012, a total of 73,360,923 ESTs have been submitted to the National Center for Biotechnology Information (NCBI) from 2430 species. EST submission to NCBI increases considerably at a monthly rate of approximately one million hits.

EST-SSRs are highly transferable for detecting the gene-rich areas within the genome. We can utilize these markers to evaluate marker transferability across taxa and conduct comparative mapping and gene functional diversity analysis in addition to genotyping. The functional EST-SSR markers should be even more useful for developing a linkage map or tagging a

 ©FUNPEC-RP www.funpecrp.com.br

viticulturally important trait. In addition, the polymorphic EST-SSR markers are much needed for genotyping, cultivar identification, and the development of a linkage map for *Vitis* species.

Research on fruit crop EST has also been given increasing attention (Zhao et al., 2008; Li et al., 2010), with the importance of grapevine in plant genomics being well reflected from grapevine EST projects that were initiated in different countries worldwide. In 2001, there were fewer than 400 ESTs from *V. vinifera* L. that were deposited in GenBank (Moser et al., 2005), but this number rose rapidly to 195,434 by July 1, 2006 (Peng et al., 2007) and 446,664 by July 1, 2012. Establishing sets of ESTs from different cultivars is important for molecular genetics and genomics because some nucleotide variations exist among cultivars.

Here, we reported the identification and characterization of 2522 unique grape EST-SSRs that were derived from a total of 362,375 grape ESTs. With this background knowledge, the objective of this study was to assess the polymorphisms among 94 F_1 hybrids from crosses between "Early Rose" and "Red Globe" using 25 EST-derived and 25 non-EST SSR markers.

MATERIAL AND METHODS

Plant material

For polymerase chain reaction (PCR) amplification and polymorphism analysis, 2 parents with 94 F_1 population (Early Rose and Red Globe) were collected from the Zhengzhou Fruit Research Institute, Chinese Academy of Agriculture Science, and which were used as the mapping population. Young fresh leaf samples were collected and frozen in liquid nitrogen and samples were stored at -40°C until use. Genomic DNA was extracted from young fresh leaves of these grape cultivars using a modified cetyltrimethylammonium bromide (CTAB) protocol (Qu et al., 1996).

Grape EST retrieval from NCBI and analysis

All grape EST that were available in the NCBI database on November 21, 2010 were retrieved. Among the total 362,375 ESTs, 2522 SSRs were identified from *V. vinifera* L. For the vector sequences, low-quality and redundant sequences were rejected with cTrans (http://www.njau.edu.cn/down/ctrans/, Xu et al., 2007) and cap3 (http://seq.cs.iastate.edu/cap3.html, Huang and Madan, 1999) softwares.

Computer programs for mining SSRs from ESTs

A Perl script program named Microsatellite (MISA) that was developed by Thiel et al., 2003 (http://pgrc.ipk-gatersleben.de/misa) was used to identify EST-SSRs. The SSRs are between 2 and 6 nucleotides in size. The minimal length of SSR was defined as 2 x 9 = 18 bp for dinucleotides, 3 x 6 = 18 bp for trinucleotides, 4 x 5 = 20 bp for tetranucleotides, 5 x 4 = 20 bp for pentanucleotides, and 6 x 3 = 18 bp for hexanucleotides. ESTs containing SSRs were assembled in Seqencher® version 4.2 (Gencodes, Ann Arbor, MI, USA) under criteria of 40% minimum overlap and 90% minimum match percentage. Based on the gene annotation number within the primer position on chromosome non-EST SSRs were found from EST. For gene annotations, we used the grape genome browser (http://www.genoscope.cns.fr/externe/GenomeBrowser/Vitis).

©FUNPEC-RP www.funpecrp.com.br

PCR amplification and verification of genomic DNA

Twenty-five pairs of grapevine EST-SSRs and 25 pairs of non-EST SSRs were used to conduct PCR amplification. PCR amplification was carried out in a 20-μL reaction system containing 2 μL genomic DNA (30 ng/μL), 0.8 μL 10 pmol of each primer, 0.1 μL Taq DNA polymerase (5 U/μL), 2 μL 10X buffer, 1.6 μL 25 mM $MgCl_2$, and 1.2 μL 2.5 mM dNTPs. The amplification of the reaction was performed in an Eppendorf Authorized Thermal Cycler using the following temperature cycling parameters: initial denaturation for 5 min at 94°C; 35 cycles of denaturation at 94°C for 40 s, the corresponding annealing temperature for 40 s, and extension at 72°C for 1 min; and a final extension step at 72°C for 10 min. PCR products were resolved by non-denaturing polyacrylamide gel electrophoresis to check the DNA banding patterns.

Data collection and analysis

In order to analyze the polymorphisms of the 2 parental grapevine lines, EST-derived and non-EST SSR polymorphic bands were visually scored as either present (1) or absent (0) and were used to create a binary data set, in which only clear unambiguous bands on non-denaturing polyacrylamide gels were chosen and scored. Data were entered in Microsoft® Excel (Microsoft Corp.) spreadsheets.

To measure the marker polymorphism, the polymorphism information content (PIC) for each EST-derived and non-EST SSR was calculated according to the formula $PIC = 1 - \sum pi^2$, where pi is the frequency of the ith allele for each SSR marker locus in the set of 94 F_1 hybrids from the cross between "Early Rose" and "Red Globe" (Weir, 1990). The PIC parameter was estimated using the PowerMarker V3.25 software (Liu and Muse, 2005).

RESULTS AND DISCUSSION

Identification and characterization of grape EST-derived and non-EST SSRs

A total of 2522 of 362,375 grapevine ESTs that were retrieved from NCBI on November 21, 2010 contained SSRs. Because some of them had multiple SSR sites, a total of 1984 SSR motifs were identified among these 2522 EST. Among the EST-derived and non-EST SSR repeats, trinucleotide repeats, which accounted for 34.09% of total SSRs, were the most abundant repeat unit followed by tetranucleotide (28.58%), dinucleotide (19.07%), pentanucleotide (12.64%), and hexanucleotide repeats (5.59%; Table 1). These findings agree with previous observations of SSR units in barley, maize, rice, sorghum, and wheat (Kantety et al., 2002). Among the SSRs, the most abundant dinucleotide repeat was AG/CT, which accounted for 85.65% of total EST-SSRs, and the most common EST-derived trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide repeats were AAG/CTT (32.55%), AAAG/CTTT (29.21%), AGAGG/CCTCT (24.45%), and AGGGGG/CCCCCT (14.18%), respectively (Table 1).

Table 1. Characterization of 2522 grape EST-SSRs.

Unit size	No. of EST-SSRs	Percentage	Abundant type	Percentage
Dinucleotide	481	19.07	AG/CT	85.65
Trinucleotide	860	34.09	AAG/CTT	32.55
Tetranucleotide	721	28.58	AAG/CTTT	29.21
Pentanucleotide	319	12.64	AGAGG/CCTCT	24.45
Hexanucleotide	141	5.59	AGGGGG/CCCCCT	14.18

Comparison of EST-derived and non-EST SSRs

A total 131 alleles were amplified using 50 primer pairs. Among them 78 were amplified from EST-derived SSRs and 53 were from non-EST SSRs. DNA polymorphisms within and/or between the grape F_1 population from the cross between "Early Rose" and "Red Globe" varieties were investigated on the basis of EST-derived and non-EST SSR markers, and polymorphisms were observed based on allele frequencies at each locus examined. The number of alleles per locus at EST-derived and non-EST-SSR ranged from 2 to 6 and 2 to 5, respectively, with an average of 3.12 for EST-derived and 2.12 for non-EST SSRs, which is comparable to the polymorphisms at SSR loci that were reported in maize (2 to 13, with an average of 6.5; Labate et al., 2003), tea (2 to 7, with an average of 4.39; Ma et al., 2010), and cucumber (2 to 8, with an average of 3.44; Mu et al., 2008). The EST-derived SSR markers showed a maximum PIC value of 1 and a minimum PIC value of 0.33. The non-EST SSR markers showed a maximum PIC value of 1 and a minimum PIC value of 0.25. The average PIC value for EST-derived SSR markers was 0.56 while that for non-EST SSR markers was 0.45.

Gene discovery is one of the most important tasks in the subsequent analysis of genome sequencing projects. ESTs are a short sub-sequence of a cDNA sequence that also represents portions of expressed genes. ESTs can be mapped in the chromosome sequences, and we investigated the sequencing project quality of grapevine by mapping 205 primer pairs from 2522 of 362,375 ESTs with each chromosome (Table 2).

Table 2. Number of EST-derived and non EST-SSR sequences located on different chromosomes.

Chromosome No.	Accession No.	Length of chromosome (bp)	Quantity of EST-derived SSR	Quantity of non-EST-SSR
chr1	NC_012007	15,630,816	7	0
chr2	NC_012008	17,603,400	8	0
chr3	NC_012009	10,186,927	12	1
chr4	NC_012010	19,293,076	10	2
chr5	NC_012011	23,428,299	13	2
chr6	NC_012012	24,148,918	11	0
chr7	NC_012013	15,233,747	11	1
chr8	NC_012014	21,557,227	7	3
chr9	NC_012015	16,532,244	5	3
chr10	NC_012016	9,647,040	6	1
chr11	NC_012017	13,936,303	8	1
chr12	NC_012018	18,540,817	11	2
chr13	NC_012019	15,191,948	8	0
chr14	NC_012020	19,480,434	15	2
chr15	NC_012021	7,693,613	6	1
chr16	NC_012022	8,158,851	5	0
chr17	NC_012023	13,059,092	11	0
chr18	NC_012024	19,691,255	13	1
chr19	NC_012025	14,071,813	8	2
Chr unknown			0	3+4 (not amplified)
Total			176	29

 ©FUNPEC-RP www.funpecrp.com.br

In this study, we used 25 EST-derived primer pairs that predict the gene within the primer position on chromosomes, but non-EST SSR primer pairs could not predict any gene within the primer position on chromosomes (Tables 3 and 4). We also made an attempt to use the EST-SSR and non-EST SSR markers to predict the gene information within 0.1 Mb of the forward and reverse primer positions on the chromosome. A total of 440 genes were found in different positions on different chromosomes using 25 primer pairs of EST-derived SSR markers (**Table S1**), and 329 genes were found using 25 primer pairs of non-EST SSR markers (**Table S2**). This may be because EST-SSRs are expressed sequences in the grapevine genome, which may be functionally associated with components of different traits, whereas the non-EST SSRs may be randomly distributed across the genome. Studies carried out in sugarcane (da Silva, 2001) and wheat (Eujayl et al., 2002) indicated that EST-SSRs were highly useful because of their high polymorphism, cross-transferability across species, and, most importantly, their association with sequences coding for function. They are found in different regions on chromosomes, such as the protein-coding and non-protein-coding sequences.

EST-derived and non-EST SSR marker development and validation

With the availability of large numbers of ESTs, the development of SSR markers from ESTs through data mining has become an efficient option for many plant species, which is also a successful way to utilize the ESTs that were released publicly. In this study, 205 unique SSR primer pairs were randomly selected, and among the 205 primer pairs, 176 EST-derived and 29 non-EST SSR primer pairs were identified (Table 5). Of the 176 EST-derived SSR primers, 25 pairs were randomly selected, all 25 primer pairs (100%) amplified the anticipated PCR products, and 21 primer pairs (84%) showed polymorphic bands (Figure 1). On the other hand, among the 29 non-EST SSR primer pairs, 25 pairs (86.20%) amplified anticipated PCR products, and 12 primer pairs (48%) showed polymorphic bands (Figure 1). This result indicated that EST-derived SSRs showed higher levels of polymorphism than non-EST SSR markers. Compared to genome-derived markers, EST-SSRs are highly transferable for detecting gene-rich areas within the genome. We can use these markers to evaluate marker transferability across taxa and conduct analysis in comparative mapping and gene functional diversity analysis, in addition to genotyping. In conclusion, large-scale EST information was generated, which can be of great use in further research on genotyping, cultivar identification, and linkage map analysis of *V. vinifera*.

Supplementary material

Table 3. Twenty-five pairs of EST-derived SSR markers and their polymorphic information content.

Marker ID	NCBI GI No.	Primer sequences (5'-3')*	Primer position on chromosome	Chromosome No.	Total alleles	Polymorphic alleles	Polymorphic information content	Gene annotation No. within the primer position on chromosome
EM002	gi:161717677	F: GGAAGCAGAAACAGCAGAGG R: GGTGGTGTGCGGATAGACTT	4417599-4417618 4417891-4417872	Chr5	3	2	0.66	GSVIVT01017 890001
EM010	gi:161721396	F: ACCGCTTCTTTGCCTCTTCT R: GATAAACCCCCTCCAGCAAT	8713612-8713631 8713911-8713892	Chr18	3	2	0.66	GSVIVT01009 478001
EM023	gi:161718390	F: CAGAAGCCCAAGAAAGATCG R: CTTCTTTGGAGCTGGTGGAC	21592915-21592934 21593075-21593056	Chr8	5	2	0.40	GSVIVT01033 314001
EM030	gi:161717492	F: GACCATGTTCTCTCCGCTTC R: CGGATGTACTCGTCCTCCAT	1263697-1263678 1263490-263509	Chr2	2	2	1.00	GSVIVT01019 517001
EM037	Contig 1178	F: CATTCCGCCATTTCAAGATT R: TAGGGTTGCCATTCTTCACC	18561748-18561729 18561589-18561608	Chr13	4	3	0.75	GSVIVT01036 582001
EM045	Contig 130	F: GACGTGGCGCTTCCTACTAC R: CACAGCCATCAATCTCTCTCC	1722810-1722791 1722608-1722628	Chr14	5	2	0.40	GSVIVT01031 147001
EM053	Contig 768	F: GCGATATGAGCCAAGACCAT R: CTGTGGAGGTTGAGGGTGAT	4306581-4306600 4306747-4306728	Chr3	1	0	0.00	GSVIVT01031 779001
EM066	Contig 293	F: AGCTTGAATCCTGGGAACCT R: TACATCCTGCTTTGGCAGTG	15949408-15949427 15949731-15949712	Chr13	1	0	0.00	GSVIVT01027 355001
EM080	Contig 968	F: TCCTCGACTACCGCAGCTAT R: CACACGGTTTGTATCGCTTG	7068516-7068497 7068245- 068264	Chr17	1	0	0.00	GSVIVT01007 970001
EM100	Contig 1394	F: TCGGCTTCACACTCCTCTCT R: GGAACCCACTTTTCCTCCTC	21714984-21714965 21714799-21714818	Chr8	1	0	0.00	GSVIVT01033 299001
EM119	gi:110732353	F: TGGAAGCGAGAATGTCAATG R: GGCACACTTGCTTAGGCTCT	21316360-21316341 21316154-21316173	Chr4	6	5	0.83	GSVIVT01026 588001
EM127	gi:110732806	F: GACCATGTTCTCTCCGCTTC R: CGGATGTACTCGTCCTCCAT	1263697-1263678 1263490-1263509	Chr2	3	2	0.66	GSVIVT01019 517001
EM130	gi:110732828	F: CCAATGAGGGCAGCAATAAC R: TCAGGAACAACGCACTCAAC	3100081-3100062 3099814-3099833	Chr17	5	3	0.60	GSVIVT01008 343001
EM137	gi:110733208	F: CGAGCCCATCTACTCACCTC R: TGTGCCGCTCCTTCTATTCT	3751261-3751280 3751433-3751414	Chr17	3	2	0.66	GSVIVT01008 273001
EM139	gi:111125110	F: AGGGAGATTGGTGGAGGTTT R: TCGGTTTCTCTGGAAAATGG	16884639-16884620 16884402-16884421	Chr11	2	1	0.50	GSVIVT01010 855001
EM150	gi:122689074	F: GGATGAAGGGCAACACATCT R: GAACCAATCAACCGAGCATT	4703290-4703271 4702955-4702974	Chr5	3	2	0.66	GSVIVT01017 920001
EM155	gi:122689350	F: GGTGTGGAGTGTTGGGAGAT R: TGGTCGCAAGTGCAACTTAT	8026360-8026379 8026566-8026547	Chr5	2	1	0.50	GSVIVT01027 809001
EM157	gi:122689538	F: CTCTGGACAACAACCCATCC R: GGAGGTGCAGAACAAGAAGC	11385202-11385221 11385460-11385441	Chr4	2	1	0.50	GSVIVT01035 252001
EM164	gi:122689756	F: CTTCTTCAGGGCACCATAGC R: CAAACCTCGACGTCTCCAAT	4242887-4242868 4242694-4242713	Chr12	4	2	0.50	GSVIVT01020 566001
EM176	gi:122690179	F: CAACGTCTCCCTTGCTTCTC R: TCCACACTCTGATTCGTTGC	5000563-5000582 5000714-5000695	Chr18	4	3	0.75	GSVIVT01009 096001
EM182	gi:122690385	F: CAAGAAGCTCCAAACCAAGC R: CGGCGACTTTCAAAGAGAAC	3394634-3394615 3394398-3394417	Chr7	3	2	0.66	GSVIVT01028 044001
G9	grffca0_001748	F: ATGGTCGTGGAATGTGTGAA R: CAATGCCTTGTGCTTGAAGA	8037601-8037582 8037430-8037449	Chr14	4	2	0.50	GSVIVT01036 25001
G14	grffca0_003143	F: TCTCTGTAATTCCCTCGATTTTT R: GAGAATCCGCCTGTTTTGAG	837689-837878 837878-837859	Chr5	3	1	0.33	GSVIVT01035 005001
G23	Contig754	F: GGAATCTTTTCCTGTTCTCA R: CCATGGTGGTGAAGATTGAA	6001035-6001016 6000832-6000851	Chr3	5	2	0.40	GSVIVT01003 172001
G32	Contig875	F: GAAGAATCCAAATGGGAGC R: GCCAATACCGTCCTTGAAGA	17979716-17979735 17980047-17980028	Chr16	3	2	0.66	GSVIVT01028 868001
Total					78	44		
Average					3.12	1.76	0.56	

*Every two primers belong to one pair. F and R = forward and reverse primers, respectively.

©FUNPEC-RP www.funpecrp.com.br

Table 4. Twenty-five pairs of non EST-SSR markers and their polymorphic information content.

Marker ID	NCBI GI No.	Primer sequences (5'-3')*	Primer position on chromosome	Chromosome No.	Total alleles	Polymorphic alleles	Polymorphic information content	Gene annotation No. within the primer position on chromosome
EM003	gi:161721399	F: TTTTCTCGTCTTGGGGTCTG R: ACTGTTCGGAGGTTGACGAC	7640568-7640549 7640246-7640265	Chr14	3	3	1	No gene found
EM020	gi:254915138	F: GAGATGGCTGTGGGATCATT R: TGCCTTTTCCTTGCACTTTT	16969702-16969683 16969547- 16969566	Chr Un	5	3	0.66	No gene found
EM068	Contig 181	F: ATTGAAGGAGCCATGGTGAG R: TGGAGGTTATGTGCCCTCTC	19762338- 19762319 19762142- 19762161	Chr7	1	0	0.0	No gene found
EM069	Contig 728	F: AAGCCGAATCCCATAGTCCT R: ACTTCCCGAACTGACCAATG	17891944- 17891925 17891747- 17891766	Chr12	3	2	0.66	No gene found
EM070	Contig 733	F: AGGCCCCACCACTTTATACC R: CCTCCCCTCAAAACCTTCTC	19213384- 19213403 19213630- 19213611	Chr un	2	1	0.50	No gene found
EM073	Contig 874	F: TGGATGGGGCTACGAATTAC R: ATGGAGGAATGACGCAGAAC	6645501- 6645520 6645802- 6645520	Chr4	4	3	0.75	No gene found
EM077	Contig 904	F: AAGTCGTGCCAACAATGGAT R: AGTTGGCAGCTGCTGATTTT	19263563- 19263582 19263796-19263777	Chr3	2	1	0.50	No gene found
EM088	Contig 1142	F: ACAACAGCCCAATCGTCTTC R: GCAGTTCCACCACCTCCTAC	1142699- 1142718 1143041- 1143022	Chr9	1	0	0.0	No gene found
EM103	Contig 1473	F: ACGCTACATGCACCTCACTG R: AGCAGCCCTGTTGGTACATC	24687393- 24687412 24674156- 24674137	Chr14	1	0	0.0	No gene found
EM105	gi:110731918	F: TACAACCCCTTCTCCTGTGG R: CTTCTGGTCCGACCTCTCAG	3991193- 3991481 3991481- 3991462	Chr5	3	2	0.66	No gene found
EM111	gi:110732057	F: TGAAGTTGACGGTGAGTTCG R: TCATGATCTGAGTGCCCAAG	2861042- 2861061 2861282- 2861263	Chr19	1	0	0.0	No gene found
EM124	gi:110732585	F: TCGCAGCAGGAGGTTAAGAG R: CCTCAATCCACGAATCCAAT	12984173- 12984192 12984382- 12984363	Chr8	1	0	0.0	No gene found
EM125	gi:110732661	F: CTGCTGCAAATTTGTGCTGT R: TCTGGACGTAACCCACATGA	5125497- 5125478 5125332- 5125349	Chr4	1	0	0.0	No gene found
EM126	gi:110732702	F: GCAGTTGGCCATTACTTGGT R: AGGGAAATACAGCCCAGGTT	25648430- 25648449 25648591- 25648573	Chr un	3	2	0.66	No gene found
EM142	gi:110733143	F: TCAGGTACGACCCTCTCAGC R: CGAGAATTCCCGCACATAGT	6567346- 6567365 6567650- 6567631	Chr11	2	1	0.50	No gene found
EM156	gi:122689463	F: ATCCACCCATTCCTTCCTTC R: TTTCGGGGTATTCCTGTGAG	20430437- 20430456 20430613- 20430594	Chr7	1	0	0.0	No gene found
EM158	gi:122689549	F: TAAAAGGCTCCTGCACCAAC R: GCTGTGCACTTTCCCAAAAT	6456799-6456818 6457033-6457014	Chr19	1	0	0.0	No gene found

Continued on next page

Table 4. Continued.

Marker ID	NCBI GI No.	Primer sequences (5'-3')*	Primer position on chromosome	Chromosome No.	Total alleles	Polymorphic alleles	Polymorphic information content	Gene annotation No. within the primer position on chromosome
EM159	gi:122689571	F: GAACATCGCGGGATACAAGT R: CAGCGCCTCCAGTTTTAGAC	9068375- 9068394 9068712- 9068693	Chr8	1	0	0.0	No gene found
EM163	gi:122689751	F: GCCACCAAACAAGCCATATC R: GTGCGGGATAGTGGTGACTT	16916680- 16916661 16916408- 16916427	Chr8	1	0	0.0	No gene found
EM165	gi:122689769	F: CTTCTCCCAAGCAATGAAGC R: AATCTTTGAGTGCCGGAATG	6320881-6320862 6320661-6320680	Chr18	1	0	0.0	No gene found
EM166	gi:122689784	F: GCAAATTGTTTCCGCAAAGT R: GCATTTAACATTAAGGGCCTGT	1285674-1285655 1285374-1285395	Chr9	3	2	0.66	No gene found
EM167	gi:122689784	F: GCAAATTGTTTCCGCAAAGT R: GCATTTAACATTAAGGGCCTGT	1285674-1285655 1285374-1285395	Chr9	1	0	0.0	No gene found
EM180	gi:122690241	F: TTGTGCCTCAATCCATTGTC R: TCCTTGGAAAAATTCCCTCT	12139039- 12139058 12139274- 12139255	Chr15	1	0	0.0	
G11	grffca0_002014	F: AGGCTGCCAGTTAGGCTTTT R: GAAACTGCGGATTCGAAGAG	22084925- 22084944 22085106- 22085087	Chr12	4	1	0.25	No gene found
G12	grffca0_002487	F: AGCAGCAGAAGAAGCAGCTC R: TCTGTTCACAAAGGGAAGCA	5845286-5845305 5845439-5845420	Chr5	4	3	0.75	No gene found
Total					53	24		
Average					2.12	0.96	0.45	

*Every two primers belong to one pair. F and R = forward and reverse primers, respectively.

Table 5. Origin of the grape EST-derived and non EST-SSR marker used for polymorphism.

Total markers	Total EST-derived SSR	Total non EST-SSR	Reference
22	20	2	Wang et al., 2012
183	156	27	Kayesh et al., 2013

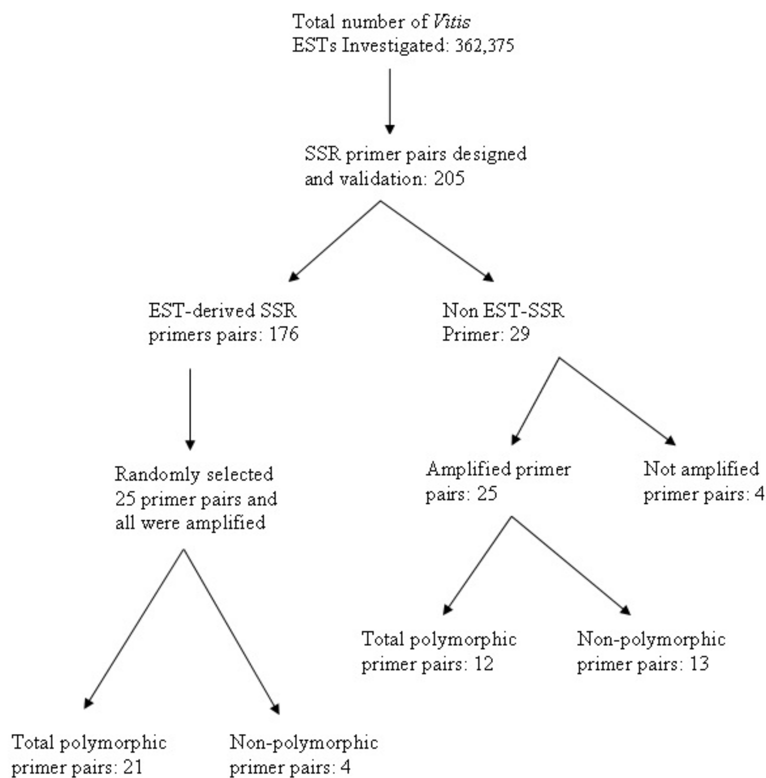

Figure 1. Flow chart of *Vitis* EST-derived and non-EST-SSR characterization.

REFERENCES

Adams MD, Kelley JM, Gocayne JD, Dubnick M, et al. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252: 1651-1656.

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

Chen L, Zhao LP and Gao QK (2005). Generation and analysis of expressed sequence tags from the tender shoots cDNA library of tea plant (*Camellia sinensis*). *Plant Sci.* 168: 359-363.

Cordeiro GM, Casu R, McIntyre CL, Manners JM, et al. (2001). Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. *Plant Sci.* 160: 1115-1123.

da Silva JAG (2001). Preliminary analysis of microsatellite markers derived from sugarcane expressed sequence tags (ESTs). *Genet. Mol. Biol.* 24: 155-159.

Depeiges A, Goubely C, Lenoir A, Cocherel S, et al. (1995). Identification of the most represented repeated motifs in *Arabidopsis thaliana* microsatellite loci. *Theor. Appl. Genet.* 91: 160-168.

Eujayl I, Sorrells ME, Baum M, Wolters P, et al. (2002). Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. *Theor. Appl. Genet.* 104: 399-407.

 ©FUNPEC-RP www.funpecrp.com.br

Feng SP, Li WG, Huang HS, Wang JY, et al. (2009). Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). *Mol. Breed.* 23: 85-97.

Gao L, Tang J, Li H, Jia J, et al. (2003). Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Mol. Breed.* 12: 261.

Huang X and Madan A (1999). CAP3: A DNA sequence assembly program. *Genome Res*. 9: 868-877.

Kantety RV, La Rota M, Matthews DE and 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.

Kayesh E, Zhang YY, Liu GS, Bilkish N, et al. (2013). Development of highly polymorphic EST-SSR markers and segregation in F_1 hybrid population of *Vitis vinifera* L. *Genet. Mol. Res*. 12: 3871-3878.

Labate JA, Lamkey KR, Mitchell SE, Kresovich S, et al. (2003). Molecular and historical aspects of Corn Belt dent diversity. *Crop Sci.* 43: 80-91.

Li X, Shangguan L, Song C, Wang C, et al. (2010). Analysis of expressed sequence tags from *Prunus mume* flower and fruit and development of simple sequence repeat markers. *BMC Genet.* 11: 66.

Lima L, Gramacho K, Gesteira A, Lopes U, et al. (2008). Characterization of microsatellites from cacao - *Moniliophthora perniciosa* interaction expressed sequence tags. *Mol. Breed.* 22: 315-318.

Liu K and Muse SV (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21: 2128-2129.

Ma JQ, Zhou YH, Ma CL, Yao MZ, et al. (2010). Identification and characterization of 74 novel polymorphic EST-SSR markers in the tea plant, *Camellia sinensis* (Theaceae). *Am. J. Bot.* 97: e153-e156.

Matsuoka K, Demura T, Galis I, Horiguchi T, et al. (2004). A comprehensive gene expression analysis toward the understanding of growth and differentiation of tobacco BY-2 cells. *Plant Cell Physiol.* 45: 1280-1289.

Moser C, Segala C, Fontana P, Salakhudtinov I, et al. (2005). Comparative analysis of expressed sequence tags from different organs of *Vitis vinifera* L. *Funct. Integr. Genomics* 5: 208-217.

Mu SQ, Gu XF, Zhang SP, Wang XW, et al. (2008). Genetic diversity of Cucumber (*Cucumis sativus* L.) germplasm by SSR. *Acta Hortic. Sin.* 35: 1323-1330.

Park JS, Kim JB, Haha BS, Kim KH, et al. (2004). EST analysis of genes involved in secondary metabolism in *Camellia sinensis* (tea) using suppression subtractive hybridization. *Plant Sci.* 166: 953-961.

Pashley CH, Ellis JR, McCauley DE and Burke JM (2006). EST databases as a source for molecular markers: lessons from *Helianthus*. *J. Hered.* 97: 381-388.

Peng FY, Reid KE, Liao N, Schlosser J, et al. (2007). Generation of ESTs in *Vitis vinifera* wine grape (Cabernet Sauvignon) and table grape (Muscat Hamburg) and discovery of new candidate genes with potential roles in berry development. *Gene* 402: 40-50.

Powell W, Machray GC and Provan J (1996). Polymorphism revealed by simple sequence repeats. *Trends Plant Sci.* 1: 215-222.

Qu X, Lu J and Lamikanra O (1996). Genetic diversity in Muscadine and American bunch grapes based on randomly amplified polymorphic DNA (RAPD) analysis. *J. Am. Soc. Hort. Sci.* 121: 1020-1023.

Remy I and Michnick SW (2004). A cDNA library functional screening strategy based on fluorescent protein complementation assays to identify novel components of signaling pathways. *Methods* 32: 381-388.

Sugui JA and Deising HB (2002). Isolation of infection-specific sequence tags expressed during early stages of maize anthracnose disease development. *Mol. Plant Pathol.* 3: 197-203.

Thiel T, Michalek W, Varshney RK and 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.

Urbanczyk-Wochniak E and Sumner LW (2007). MedicCyc: a biochemical pathway database for *Medicago truncatula*. *Bioinformatics* 23: 1418-1423.

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

Wang XC, Guo L, Shangguan LF, Wang C, et al. (2012). Analysis of expressed sequence tags from grapevine flower and fruit and development of simple sequence repeat markers. *Mol. Biol. Rep.* 39: 6825-6834.

Weir BS (1990). Genetic Data Analysis - Methods for Discrete Population Genetics Data. Sinauer Associates Inc., Sunderland.

Xu H, Yang L, Xu P, Tao Y, et al. (2007). cTrans: generating polypeptide databases from cDNA sequences. *Proteomics* 7: 177-179.

Yamada-Akiyama H, Akiyama Y, Ebina M, Xu Q, et al. (2009). Analysis of expressed sequence tags in apomictic guineagrass (*Panicum maximum*). *J. Plant Physiol.* 166: 750-761.

Zhao L, Ma C and Chen L (2008). Construction and expressed sequence tags analysis of young roots cDNA library of tea plant. *Mol. Plant Breed.* 6: 893-898.

Zhao WW, Zhong C, Cao H, Chen L, et al. (2009). Study on methods of extracting total RNA from grape inflorescence. *J. Anhui Agric. Sci.* 37: 16161-16162.

 ©FUNPEC-RP www.funpecrp.com.br