African Journal of Biotechnology Vol. 11(6), pp. 1337-1363, 19 January, 2012
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB11.1509
ISSN 1684–5315 © 2012 Academic Journals

Full Length Research Paper

Novel expressed sequence tag- simple sequence repeats (EST-SSR) markers characterized by new bioinformatic criteria reveal high genetic similarity in sugarcane (*Saccharum* spp.) breeding lines

Kittipat Ukoskit[1]*, Penjun Thipmongkolcharoen[1] and Prasert Chatwachirawong[2]

[1]Department of Biotechnology, Faculty of Science and Technology, Thammasat University (Rangsit Campus) Klong Luang, Pathumtani 12121, Thailand.
[2]Department of Agronomy, Faculty of Agriculture, Kasetsart University, Kamphaengsean Campus, Kamphaengsean, Nakhon Pathom 73140, Thailand.

Accepted 21 September, 2011

Using different bioinformatic criteria, the SUCEST database was used to mine for simple sequence repeat (SSR) markers. Among 42,189 clusters, 1,425 expressed sequence tag- simple sequence repeats (EST-SSRs) were identified *in silico*. Trinucleotide repeats were the most abundant SSRs detected. Of 212 primer pairs selected, based on repeat patterns of n≥8 for di-, tri-, tetra- and penta-nucleotide repeat motifs verified using 15 sugarcane (*Saccharum* spp.) genotypes and marker segregation using F1 progenies of a cultivated sugarcane and *Saccharum spontaneum*, 191 loci were identified. All new EST-SSR loci detected a total of 1,529 markers ranging from 2 to 21 markers per locus, with an average of eight markers per locus. Observed polymorphism ranged from 0.12 to 0.93 with a mean of 0.74. A total of 426 and 333 markers were putatively identified as simplex in the cultivated sugarcane and *S. spontaneum*, respectively and corresponding to 2.23 and 1.74 markers per primer, respectively while 167 markers were identified as double-simplex markers, with 0.87 markers per primer. Cluster analysis revealed a high genetic similarity among the sugarcane (*Saccharum* spp.) breeding lines which could reduce the genetic gain in sugarcane breeding.

Key words: sugarcane, expressed sequence tag- simple sequence repeats (EST-SSRs) markers, genetic similarity.

INTRODUCTION

In current years, with the rapid increase of expressed sequence tag (EST) sequence in public data base, the development of EST containing simple sequence repeats (SSRs) becomes an attractive choice for the development of SSR markers. With evolving bioinformatic tools, it is easy to download the ESTs from public databases, scan for EST containing SSRs and develop EST–SSR markers at a large scale with a time saving and cost effective way (Kantety et al., 2002; Yan et al., 2008). EST-SSRs have some advantages over genomic SSRs since these markers are derived from expressed genes, they are more conserved and have a better potential for their applications across the species than anonymous DNA markers like amplified fragment length polymorphisms (AFLPs), random amplified polymorphic DNA (RAPD) and genomic SSR markers (Feng et al., 2009). Consequently, they are useful as anchor markers for identifying conserved genomic regions among species and genera, comparative genomics, and evolutionary studies (Cordeiro et al., 2001; Kantety et al., 2002; Thiel et al., 2003). Moreover, EST-SSRs may be directly

*Corresponding author. E-mail: ku@tu.ac.th.Tel: 662- 5644500. Fax: 662-5644500.

related with a coding gene, used as functional SSR markers, and are better resources for their use in breeding since they may be directly associated with the genes affecting a particular trait (Andersen and Lubberstedt, 2003).

The sugarcane EST project (SUCEST) generated 291,689 ESTs (Vettore et al., 2001). Trimming and clustering ESTs were performed to obtain 43,141 clusters (Telles and Silva, 2001). These large datasets of non-redundant (NR) EST sequences are publicly released on the public domain sugarcane EST (http://sucest.lbi.ic.unicamp.br/public/) and available for developing EST–SSR markers at a large scale. Although, the sugarcane EST database had been surveyed for ESTs containing SSRs (Oliveira et al., 2007; Oliveira et al., 2009; Pinto et al., 2004; Silva, 2001), a large number of EST–SSR markers are still available for marker development.

The current study was designed (1) to develop and characterize a novel set of EST-SSR markers from the sugarcane EST using different search EST-SSR criteria from previous reports (2) to evaluate the ability to detect polymorphism and to determine simplex alleles segregating in mapping population derived from an interspecific hybridization between a cultivated sugarcane and *Saccharum spontaneum* (3) to assess their potential for diversity and pedigree relationship studies. The new EST-SSR markers developed from this study provide a robust set of polymerase chain reaction (PCR) primers that are a useful addition for functional sugarcane genome mapping facilitating introgression of favorable wild alleles in sugarcane breeding programs and for sugarcane genotyping.

MATERIALS AND METHODS

Plant materials

Thirteen sugarcane breeding lines and 2 cultivars, Phil66-107 and KU60, (Table 1) were used for evaluation of selected EST-SSR markers. Ten random F_1 progenies obtained from the interspecific cross between the cultivated sugarcane 'Phil6607' and *S. spontaneum* 'S6' were used for simplex marker segregation analysis. Leaf material was collected and DNA was extracted following the cetyltrimethylammonium bromide (CTAB) method described by (Gawel and Jarret, 1991). DNA quantity and quality were determined using agarose gel electrophoresis and spectrophotometric measurements, and then the samples were diluted appropriately for marker analyses.

Data mining for EST-SSR markers and primer design

We collected 42,189 clusters from 37 libraries publicly provided from the SUCEST database and surveyed electronically using the PERL program, Simple Sequence Repeat Identification Tool(SSRIT), downloaded from the Cornell University web site http://www.gramene.org/gramene/searches/ssrtool. The parameters in SSRIT program were set for detection of di-, tri-, tetra-, and penta nucleotide motifs with a minimum of 5 repeats. The identified ESTs containing repeat motifs were subjected to masking of other repeat sequences from the Gramineae family as well as low-complexity sequences, including the SSRs and interspersed repeats, with the RepeatMasker Program (A.F.A. Smit, R. Hubley & P. Green RepeatMasker at http://repeatmasker.org). The Primer3 software (freely available at http://www.genome.wi.mit.edu/genome_software/ other/ primer3.html) was used to design the primers (Rozen and Skaletsky, 2000) to amplify the selected EST-SSRs based on the following parameters: primer length from 18 to 24 with 20 as the optimum, PCR product size from 150 to 350, optimum annealing temperature of 60°C, and GC contents from 40 to 70%, with 50% as optimum. PCR reactions were carried out in 20 µl reaction volumes containing 1× Buffer, 1.5 mM of $MgCl_2$, 20 µM of each dNTPs, 0.25 µM of each primer, 0.5 U of *Taq* DNA polymerase and 20 ng of DNA template and were performed on MJ Thermal Cycler at 94°C for 5 min, 40 cycles of 30 s at 94°C, 60 s at appropriate annealing temperature of each primer and 1 min at 72°C, with a final extension of 7 min at 72°C. PCR products were run on 6% denaturing polyacrylamide gels and silver-stained as described by (Benbouza et al., 2006).

Gene identity of characterized EST-SSRs

To obtain sequence homology and putative function of genes represented by the EST-SSRs characterized from the Repeat Masker program, the EST-SSRs were BLAST searched against NCBI non-redundant (nr) database using Blast2GO freely available through Java Web Start at http://www.blast2go.de; (Conesa et al., 2005). BLASTX results were loaded into the program and the default settings were used to assign GO terms to all EST-SSRs.

Evaluation of EST-SSR polymorphisms and determining simplex alleles

Primarily, the selected EST-SSRs were screened for polymorphism between the two parents, Phil6607 and S6 used to generate a mapping population for constructing linkage maps. Secondly, the resulting polymorphic EST-SSRs evaluated between the two parents were determined for putative simplex markers (markers present in one parent but absent in another and segregate 1:1) and double simplex markers (marker present in both parents and segregate 3:1; (Grattapaglia and Sederoff, 1994; Ritter et al., 1990) using 10 random progenies of F_1 mapping population. Thirdly, based on high polymorphism between the two parents and homology to genes of interest, a subset of 212 primer pairs were selected to determine the polymorphism information content (PIC; Botstein et al., 1980) in 15 sugarcane genotypes, using the equation.

$$PIC = 1 - \sum_{i=1}^{k} p_i^2 - \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} 2 p_i^2 p_j^2$$

where p is the frequency of the j^{th} allele pattern of EST-SSR locus i^{th} and k is the number of allele pattern of EST-SSR locus. The selected highly polymorphic 212 EST–SSRs were tested to ascertain the genetic relationships by cluster analysis among 15 sugarcane genotypes including *S. spontaneum* used as an out group. The binary data were used to calculate Jaccard similarity coefficients (Jaccard, 1908) as indicators of genetic similarity

Table 1. Sources and parents of sugarcane clones used for evaluation of developed EST-SSR markers.

Clone	Source	Pedigree			
		Female	Breeding*	Male	Breeding
TBy20-0556	Breeding line	K83-74	OCSB	K84-200	OCSB
TBy20-0732	Breeding line	K83-74	OCSB	K84-200	OCSB
TBy20-0734	Breeding line	K83-74	OCSB	K84-200	OCSB
TBy20-0794	Breeding line	K84-200	OCSB	K92-166	OCSB
TBy20-1300	Breeding line	K83-74	OCSB	K84-200	OCSB
TBy20-1455	Breeding line	Co775	Thai Major parent	K84-200	OCSB
TBy20-2248	Breeding line	K83-74	OCSB	K84-200	OCSB
NRCT01-0398	Breeding line	K83-74	OCSB	UT1	DA
NRCT01-0663	Breeding line	KU50	KU	B4362	Thai Major parent
NRCT01-0877	Breeding line	UT1	DA	K83-74	OCSB
NRCT01-1192	Breeding line	UT1	DA	K83-74	OCSB
NRCT01-1202	Breeding line	UT1	DA	KU50	KU
NRCT01-1256	Breeding line	Co775	Thai Major parent	K82-32	OCSB
KU60-1	Cultivar	Co775	Thai Major parent	K84-200	OCSB
Phil6607	Cutivar	Phil5460	Unknown	Co440	Thai Major parent
S6	*S. spontaneum*				

*OCSB = Office of the Cane and Sugar Board; KU= Kasetsart University; DA = Department of Agriculture.

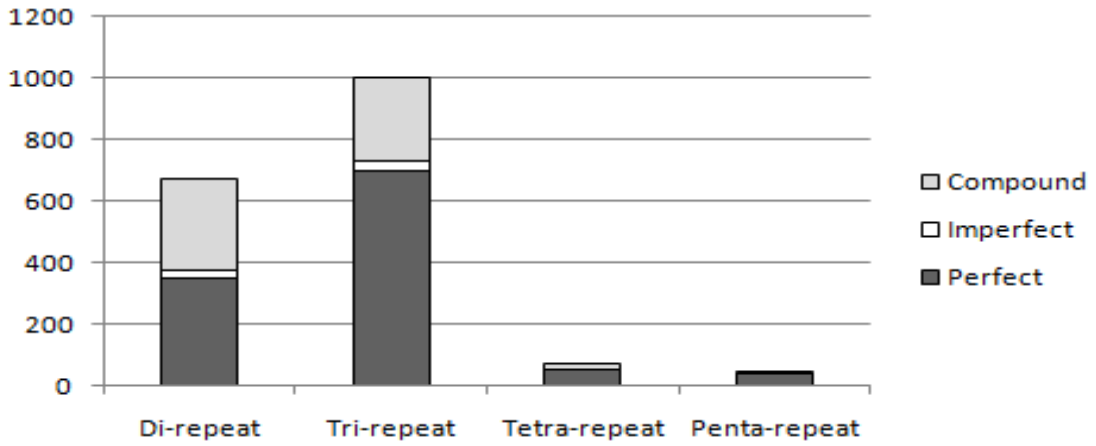

Figure 1. Frequency of repeat motif classes.

among pairs of genotypes as follows: GS*ij* = *a*/(*a*+ *b* + *c*) where GS*ij* is the genetic similarity measurement between individuals i and j, *a* is the number of polymorphic bands that are shared by *i* and *j*, *b* and *c* are the number of bands present in individual *i* and *j*, respectively. The resulting similarity matrices were used to construct dendrograms by the unweighted pair-group method with arithmetic means (UPGMA) using software packages NTSYSpc ver 2.02 (Rohlf, 1998).

RESULTS

Type and frequency of EST-SSRs

A search of the EST from the SUCEST database usingthe PERL program, SSRIT found 4,401 ESTs containing repeat motifs according to the search criteria. Since the SSRIT program identifies ESTs solely based on the fact that they contain a certain repeat motif, all these ESTs containing repeat motifs were subjected to masking of other repeat sequences from the Gramineae family as well as low-complexity sequences, including the SSRs and interspersed repeats, with the Repeat Masker pro-gram. A total of 1,425 EST-SSRs were identified, accounting for approximately 3.4% of the total cluster searched. Of the total EST-SSRs identified (Figure 1), 1,141 (80.1%) were perfect repeats representing the majority of the SSR structure types, 31

Table 2. Number and frequency of different SSR-repeat motifs identified in the analyzed 42,189 clusters from the SUCEST database.

Repeat motif	Number of repeat										Total	Frequency (%)
	5	6	7	8	9	10	11–15	16–25	26–40	41–60		
AG/CT	55	34	24	65	42	21	57	40	4	0	342	19.1
AC/GT	33	15	4	22	16	8	11	12	3	1	125	7
AT/AT	30	11	6	14	6	8	26	15	16	3	135	7.6
CG/CG	51	10	1	5	1	0	0	0	0	0	68	3.8
ACG/CGT	28	18	7	8	2	1	0	0	0	0	64	3.6
ACT/AGT	3	0	1	1	0	0	0	0	0	0	5	0.3
ATT/AAT	1	1	1	1	2	0	4	1	0	0	11	0.6
AAC/GTT	4	4	7	6	5	0	2	0	0	0	28	1.6
CAT/ATG	10	6	5	1	2	3	1	1	0	0	29	1.6
CTT/AAG	12	3	7	4	3	0	1	2	0	0	32	1.8
GGT/ACC	23	15	10	5	7	1	2	0	0	0	63	3.5
CCT/AGG	50	35	27	11	6	2	2	0	0	0	133	7.4
AGC/GCT	60	42	20	14	6	1	4	0	0	0	147	8.2
CGG/CCG	200	132	88	42	22	2	6	0	0	0	492	27.5
AAAG/CTTT	4	2	1	2	1	1	2	0	0	0	13	0.7
AAAT/ATTT	2	0	1	0	0	0	0	0	0	0	3	0.2
ACAT/ATGT	4	1	0	2	0	0	0	1	0	0	8	0.4
AGCT/AGCT	2	1	0	0	0	0	0	0	0	0	3	0.2
AGGA/TCCT	1	1	1	1	1	0	0	0	0	0	5	0.3
AGGC/GCCT	5	0	0	0	0	0	0	0	0	0	5	0.3
ATGG/CCAT	5	0	0	0	0	0	0	0	0	0	5	0.3
CAAT/ATTG	2	0	0	0	0	0	0	0	0	0	2	0.1
CCCG/CGGG	4	0	0	0	0	0	0	0	0	0	4	0.2
CCTC/GAGG	7	2	0	0	0	0	0	0	0	0	9	0.5
CGCT/AGCG	2	0	0	0	0	0	0	0	0	0	2	0.1
TCTA/TAGA	1	1	0	0	0	0	0	0	0	0	2	0.1
Other tetra-repeats	9	1	0	0	0	0	1	0	0	0	11	0.6
ATGTA/TACAT	2	0	0	0	0	0	1	0	0	0	3	0.2
ATCCA/TGGAT	2	0	0	0	0	0	0	0	0	0	2	0.1
CCTTT/AAAGG	2	0	0	0	0	0	0	0	0	0	2	0.1
TCTCT/AGAGA	0	1	0	1	0	0	0	0	0	0	2	0.1
TCTCC/GGAGA	2	1	0	0	0	0	0	0	0	0	3	0.2
CTTTT/AAAAG	3	0	0	0	0	0	0	1	0	0	4	0.2
TCCCT/AGGGA	3	1	0	0	0	0	0	0	0	0	4	0.2
Other penta-repeats	16	2	0	1	1	0	2	0	0	0	22	1.2
Total	638	340	211	206	123	48	122	73	23	4	1,788	100

(2.1%) were imperfect repeats with one or more interruptions in the run of repeats and 253 (17.6%) were compound repeats with adjacent tandem simple repeats of a different sequence, resulting in the total number of SSRs of 1,788 obtained.Analysis of SSR motifs in this study (Figure 1 and Table 2) revealed that the trinucleotide motifs were the most abundant type of SSRs found in the database searched (56.2%), followed by di- (38.3%), tetra- (4%), and penta-nucleotide repeats (2.3%). The dominance of trinucleotide motifs in this study may be explained by the suppression of non-trinucleotide motifs in coding regions because of the risk of frame shift mutations that can occur when there is length variation in these motifs (Thiel et al., 2003; Varshney et al., 2005). Among the dinucleotide motif sequences, AG/CT motif was the most common (19.1%)

followed by AT/AT (7.6%) and AC/GT (7.0%) motifs, whereas CG/CG motif was the least common (3.8%).

Among the trinucleotide motifs sequences, the motif CGG/CCG was the most common (27.5%) followed by the motifs AGC/GCT (8.2%) and CCT/AGG (7.4%) while the motif ACT/AGT was the least common (0.3%). However, the most common tetra- and pentanucleotide motif sequences were found in insig-nificant numbers (0.7 and 0.2%, respectively) due to the high cut-off used for mining of tetra- and pentanucleotide motifs.

Gene annotation and function of characterized EST-SSRs

Of 1,425 EST-SSRs, the BLAST2GO searches showed that 856 (60%) matched to genes of known functions at e values $<10^{-6}$, while 222 (16%) and 342 (24%) had matched to hypothetical proteins and had no significant matches, respectively. To provide a general repre-sentation of the annotation, the Slim GO Classification was obtained for the whole set of EST-SSRs. Of the biological processes, 260, 172, 99, 99, 53, 45 and 29 EST-SSRs were related with cellular process, metabolic process, localization, establishment of localization, biological regulation, response to stimuli and develop-mental process, respectively. In the category cellular component, two main types were associated with cellular and organelle, 461 and 293, respectively. Under the category of molecular function, the vast majority of EST-SSRs were involved in catalytic activity and in binding activities, 132 and 130 respectively. Among the known function EST-SSRs, numerous EST sequences related to a wide range of proteins of interest (Supplementary Table 1) including enzymes involved in sugar metabolism, disease resistance related proteins, abiotic related proteins and growth regulatory proteins.

EST-SSR polymorphism and segregation analysis

Of the total 1,425 EST-SSRs identified, the selected 424 and 36 primer pairs were designed based on SSRs containing repeat patterns with $n \geq 8$ and on homology to genes of interest, respectively and then assayed to detect polymorphisms between the two parents, Phill6607 and S6. Four hundred and thirty-four (434) primer pairs (94.4%) successfully amplified either Phill6607 or S6 while 26 primer pairs completely failed to yield PCR products from both parents at various annealing temperatures. Of the total 460 primer pairs, 74 (16.1%) and 26 (5.7%) primers pairs failed to amplify S6 and Phill6607, respectively. Among the 434 working primer pairs, 424 produced PCR products at the expected sizes, while 20 primer pairs yielded larger PCR product size than expected from the EST sequence, likely due to the presence of small introns. The EST-SSR markers developed in this study were highly polymorphic between the parents and among 14 sugarcane genotypes (Fig 2a). The polymorphism assay revealed that 407 primer pairs (93.7%) were polymorphic between the two parents reflecting the fact that the alleles between the cultivated sugarcane 'Phil6607' and *S. spontaneum* are distinct. The dominant scoring of the SSR bands yielded 1, 256 and 1,115 total polymorphic bands corresponding to an average of 3.1 and 2.7 polymorphic bands per primer pair in Phill6607 and S6, respectively.

All 407 polymorphic primers were tested for the simplex marker segregation in the 10 randomly selected F_1 progenies generated from the interspecific hybridization between the two parents. The simplex mrkers and double simplex markers in pseudo testcross configuration were expected to flow 1:1 (fig 2b) and 3:1 segragation patterns in the f_1 progeny, respectively. All polymorphic primers generated 2,371 polymorphic markers (bands). Of the total polymorphic markers, 828 and 165 markers were putatively identified as simplex and double simplex corresponding to 2 and 0.4 markers per primer, respectively. Based on high polymorphism between the two parents and on homology to genes of interest, a subset of 212 primer pairs were selected from 407 polymorphic primers to determine informativeness of EST-SSR markers in 15 sugarcane. Characteristics, putative function and segregation pattern of the 212 EST-SSR markers are shown in Supplementary Table 1.

Comparing redundancy to other EST-SSR data mining works in sugarcane (Oliveira et al., 2007, 2009; Pinto et al., 2006; Silva, 2001) illustrates that of 212 primer pairs, 191 primer pairs are new EST-SSR markers developed from this study (Supplementary Table 2). The new EST-SSR primer pairs detected a total of 1,529 markers ranging from 2 to 21 with the average of 8 markers per locus. The mean of PIC determined from 191 primer pairs were similar to those obtained from a previous study (Pinto et al., 2006). The PIC values ranged from 0.12 to 0.93 (except five monomorphic loci) with a mean of 0.74. 426 and 333 markers were putatively identified as simplex in Phill6607 and S6, corresponding to 2.23 and 1.74 markers per primer, respectively, while 167 markers were identified as double simplex corresponding 0.87 markers per primer.

Genetic similarity

The similarity coefficient based on 1,743 polymorphic markers (alleles) scored from 212 primer pairs ranged from 0.56 to 0.80 with a mean of 0.72. Based on the marker data, the dendrogram generated with the UPGMA clustering method illustrates that the tested germplam were clearly resolved in two distinct clusters and one separated commercial cultivar on the level of similarity

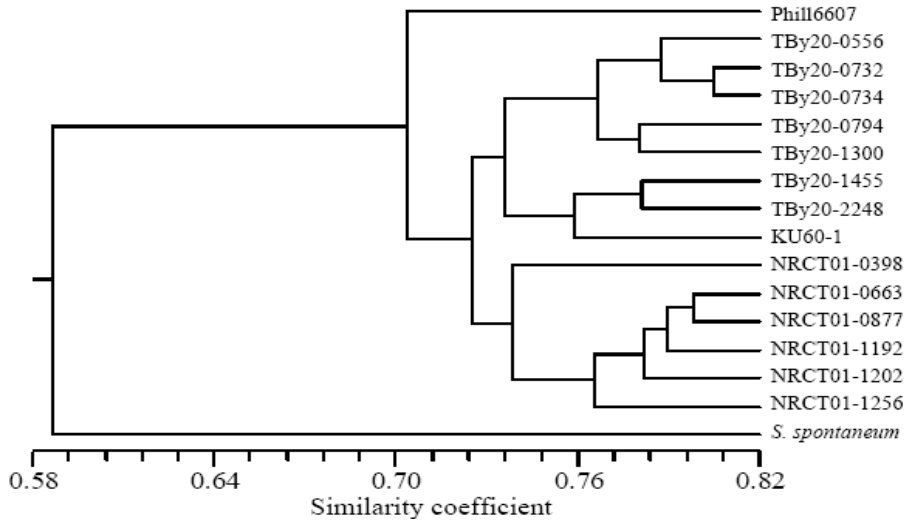

Figure 2. The UPGMA dendrogram of the 13 sugarcane breeding lines and 2 cultivars based on 212 EST-SSR primer pairs. *S. spontaneum* (S6) was included as an out species.

coefficient 0.68 as expected of their pedigree (Figure 2). *S spontanuem* were included in the analysis as an out group and clearly differentiated from the sugarcane breeding lines and two commercial cultivars. TBY family and KU60-1 cultivar was grouped on the level of similarity coefficient of 0.72, while NRCT family was grouped in the same cluster on the level of similarity coefficient of 0.71. The dendrogram further validated the placement of the old commercial cultivar, Phil6607, as the most distant to the new breeding lines and the new commercial cultivar, KU60-1.

DISCUSSION

On the basic of the average length of 790 bp per EST sequence (data from Blast2Go), at least one SSR motif was found per 18.6 kb in the approximately 33.3 Mb ESTs that were searched. Data mining of EST-SSRs in cotton and wheat showed close values with one SSR for every 20.0 (Cardle et al., 2000) and 15.6 kb (Kantety et al., 2002), respectively. This value was lower than that of barley (1/6.3 kb; Thiel et al., 2003), soybean (1/7.4 kb; Cardle et al., 2000) coffee (1/1.56 kb; Aggarwal et al., 2007) tomato (1/11.1 kb; Gupta et al., 2003) *Medicago truncatula* (1/12 kb; Mun et al., 2006). However, the density of EST-SSRs in wheat (1/9.2 kb; Cardle et al., 2000) and *M. truncatula* (1/1.7kb; Gupta and Prasad, 2009) was varied in other studies. In sugarcane, a density of one SSR every 16.9 kb was estimated from a previous study (Pinto et al., 2004). The abundance level of EST-SSR may vary because of SSR search criteria, SSR mining tools and the size of the database searched (Varshney et al., 2005). In the present study, the parameters set for detection of di-, tri-, tetra-, and penta nucleotide motifs in SSRIT program were a minimum of 5 repeats, subsequently, the identified EST-SSRs were subjected to masking for interspersed repeats and low complexity DNA sequences using the Repeat Masker program whereas surveying EST containing SSRs in sugarcane from the previous study (Cordeiro et al., 2001; Pinto et al., 2004) used the BLASTn software to search for dinucleotide, trinucleotide, and tetranucleotide repeat patterns with $n \geq 7$, 5 and 3, respectively.

Di- and tri-nucleotide repeat motifs were mostly reported in plants, but the frequency of repeat motifs were different. Di-nucleotide repeat motifs were dominant in coffee (Aggarwal et al., 2007) apricot, peach (Jung et al., 2005) and Kiwi (Fraser et al., 2004). The analysis of our results revealed that the trinucleotide motifs were the leading repeat motif type of SSRs found in the database searched. This result is cosistent with previous research in sugarcane (Duarte Filho et al., 2010), and close evolutionary species such as in maize, rice, sorghum, wheat (Kantety et al., 2002), barley (Thiel et al., 2003), and oat (Becher, 2007). Based on the previous study (Pinto et al., 2004), tetra-nucleotide repeat motifs were reported to be the most abundant in sugarcane. This also verified a previous point of view that EST-SSR frequency was related to the search criteria (Aggarwal et al., 2007; Kantety et al., 2002; Pinto et al., 2004; Thiel et al., 2003).

Duarte Filho et al. (2010, Sugar Tech Volume 12, Number 2, 145-149). The result is very similar, were di-, tri- and tetracleotide is very similar with the presents here.

Earlier reports on the abundance of different SSR motifs in plant databases showed that AT motif was the most common repeat motif type (Lagercrantz et al.,

1993). Among the dinucleotide motif sequences in our sugarcane EST-SSRs, AG/CT motif was the most common (19.1%). This contrasts to the report of Pinto et al. (2004) in which, AT/AT motif was observed preferentially among dinucleotide motif sequences. The AG/CT motif was also the most frequently observed EST-SSRs in plants (Scott et al., 2000; Kantety et al., 2002; Gao et al., 2004; Thiel et al., 2003; Saha et al., 2004; Feng et al., 2009). Expressed sequences had shown a higher frequency of AG repeats than AT repeats (Morgante et al., 2002; Mun et al., 2006). A wide variety of tri-nucleotide repeat motifs were represented at high percentages. However, the repeat motif CCG/CGG was highly represented in monocots (Cho et al., 2000; Kantety et al., 2002; Thiel et al., 2003). This result is in agreement with our results and the result from the previous surveyed in SUCEST data base (Pinto et al., 2004). Rich GC content in rice was reported in the coding regions (Cho et al., 2000). The GC content was significantly higher in monocot species than in dicot species. Codon bias had been reported to correlate with GC content at the third codon position (Kawabe and Miyashita, 2003). This could be one of the reasons why CCG/CGG motifs are present at such high frequencies in EST collections of monocot species. This result (motifs rich in GC – CCG/GGC or GCG/CGC) is similar to Duarte Filho et al. (2010), where trinucleotidie rich in CG is most frequent in SUCEST, different to Pinto et al. (2004). You can improve your discussion, including this information about EST-SSR in SUCEST.

The result of transferability of EST-SSR markers developed from the present study indicated that the rate of amplification in the commercial cultivar (94.3%) was higher than that of its wild relative, *S. spontaneum* (83.9%), for the primer pairs which were designed from ESTs generated from commercial cultivars (Vettore et al., 2001) have a high homology sequence for annealing to the DNA template of the commercial cultivar. The failure of PCR amplification in both *Saccharum* species may be explained by the fact that primers extend across a splice site or that there are large introns in the genomic sequence. In the present study, the total of 460 and the subset new 191 EST-SSR primer pairs revealed that EST–SSRs were highly polymorphic between the two parents, the cultivated sugarcane and *S. spontaneum* and showed a high PIC (0.74) among sugarcane breeding lines determined in this study. The high PIC and the average number of alleles (8.8) showed close values to previous reports in analysis of EST-SSR informativeness (Oliveira et al., 2009; Pinto et al., 2006, 2004). The high percentage of polymorphism between the two parents is due to a different genome composition and a complex polyploid with different chromosome numbers between the cultivated sugarcane, Phil6607, (2n = 10 0-130 chromosomes) and *S. spontaneum* (2*n* = 40-128 chromosomes), S6, (Price, 1963). Owing to the high ploidy number of sugarcane, most EST-SSRs yielded more than two PCR products which were assumed to be alleles. The multiple allele characteristic of SSR combined with the polyploidy nature of sugarcane, which resulted in the high PIC providing the capability of EST-SSR to create unique sugarcane fingerprints.

The new 191 EST-SSR primers generated the larger numbers of putative simplex markers (759 markers) both in Phill6607 and S6 as well as double simplex markers (167). As reflected by the much higher percentage (80%) of putative simplex markers, these types of markers represented alleles differed by the interspecific hybridization. In the case of crossing between two parents sharing alleles of heterozygous markers, the larger proportion of double simplex markers would be expected (Garcia et al., 2006). Although, double simplex markers are less informative than simplex markers (Maliepaard et al., 1998; Wu et al., 2000) double simplex markers can provide a locus connection between the mapping parents (Grattapaglia and Sederoff, 1994). Regarding both types of markers, these EST–SSR markers will be very useful for its potential to incorporate both improved and wild alleles in sugarcane linkage maps by using the simplex polymorphism approach (Wu et al., 1992) facilitating introgression favorable wild alleles with known function in sugarcane breeding programs.

The achievement of the sugarcane breeding program lies in the proper choice of parents based on genetic diversity. A cluster analysis of genetic relationships performed in sugarcane breeding lines and commercial cultivars using EST-SSR markers developed from the present study revealed two major groups of sugarcane breeding lines. This clustering result corresponded well to their known pedigree relationships and source of the breeding program. There are three main organizations in Thailand conducting research in sugarcane breeding, the Office of the Cane and Sugar Board (OCSB), the Department of Agriculture (DA), and Kasetsart University (KU), which have conducted breeding programs during the past 40 years in Thailand. Within the TBy breeding lines (including cv. KU60-1), all of them were derived from a cross between two parents from OCSB except TBy20-1455 and KU60-1 which one parent was the Thai major parent used in most breeding programs in Thailand. Therefore the clustering of TBy breeding lines and KU60-1 was due to the fact that their parental breeding lines are the same or are very close to each other in a breeding program. Although, the NRCT breeding lines involved germplasm from all three programs (Table 1), they were still grouped in the same cluster but separated from TBy breeding lines. The clustering of NRCT breeding lines was contributed by using common parents within them. Clearly, separation of the old cultivar, Phill6607, from the new breeding lines was due to dissimilarity of the parents used for breeding the cultivar, as compared to those used for generating

the new breeding lines. It should be noted that the genetic resources have been used repeatedly among the three breeding programs. Consequently, the sugarcane breeder should, in the future, carefully choose more distantly related clones for crossing in order to broaden genetic variation. The consequence of large genetic variation existing over the whole national crop would also reduce the genetic vulnerability to pests and disease of sugarcane cultivars used in commercial production. The main reason for the failure of base broadening programs is the inability to trace or follow the incorporated germplasm into the germplasm of the advanced breeding population through visual selection. EST-SSR markers developed from this study will help breeders to investigate germplasm for the selection of genetically distant parents used in future breeding, and for the selection of particular sugarcane genotypes based on higher variability among the progenies in order to broaden the gene pool in the breeding programs.

Our study also revealed a high similarity coefficient (0.70) among the two breeding line families, including Phill6607 and suggested that the *Saccharum* germplasm collection in the breeding programs presented a genetically narrow base. This result supports sveral reports of narrow genetic varition in sugarcane. Refferences to the narrow genetic base of sugarcane varieties are available from several regions, including commercial sugarcane clones from Brazil(Duarte Filho et al., 2010; Oliviera et al. 2009; Pinto et al. 2006)USA (Alwala et al. 2006; Pan 2006; Glynn et al. 2009) Mexico(Rodriguez et al. 2005) tropical and subtropical regions (Selvi al. 2003). An early report on the coefficient of parentage (Chang et al., 1991) of major sugarcane clones in Thailand revealed that the major ancestral S. *officinarum* clone in the pedigrees of the OCSB, DA, and KU cultivars was cv. Black Cheribon (Chatwachirawong and Srinives, 1999). For the *Saccharum barberi,* cv. Kansoer was the predominant genetic resource of OCSB and DA, whereas cv. Kansoer and cv. Chunee predominated at KU (http:// cropthai.ku.ac.th/coefparent/fracoef.htm). Consequently, the narrow genetic base of Thai sugarcane germplasm is descended primarily from the ancestral cultivar Black Cheribon contributing most germplasm to the improved cultivars developed by all three programs. Previous studies have also indicated that the genetic basis of modern hybrid sugarcane cultivars are essentially derivatives of no more than 15-20 nobilized cultivars (Berding and Roach, 1987; D'Hont et al., 1995; Ming et al., 2010). Since early interspecific hybridization between *Saccharum officinarum* and other species, mostly *S. spontaneum*, followed by a back-crossing process referred to as nobilization, was so successful in producing superior hybrid cultivars, they were selected and became the ancestry of most, if not all, modern cultivars grown in most countries today (Jackson, 2005). Consequently, a genetic bottleneck has occurred in development of sugarcane. Utilization of wild species with known functional markers developed from this study to tag desirable and undesirable genes in *S. spontaneum* would be an effective way to change this situation to perform more effective introgression breeding for a short- and long-term breeding program.

ACKNWOLEDGEMENT

The authors gratefully acknowlege Dr. Sittichoke Tangpatronrang for his assistance with some parts of functional analysis. This was supported by grants from NATIONAL Centre for Genetic Engineering and Bioyechnology (BIOTEC), Thailand.

REFERENCES

Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L (2007). Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. Theor. Appl. Gene. 114: 359-372.

Alkwala S, Suman A, Arro JA, Veremis JC, Kimbeng CA (2006). Target region amplification polymorphism (TRAP) for assessing genetic diversity in sugarcane germplasm collections. Crop Sci. 46:448-455

Andersen JR, Lubberstedt T (2003). Functional markers in plants. Trends in Plant Science 8: 554-560.

Becher R (2007). EST-derived microsatellites as a rich source of molecular markers for oats. Plant Breed. 126: 274-278.

Benbouza H, Jacquemin JM, Baudoin T, Mergeai G (2006). Optimization of a reliable, fast, cheap and sensitive silver staining method to detect SSR markers in polyacrylamide gels. Biol. Agron. Soc. Environ. 10: 77-81.

Berding N, Roach B (eds) (1987). Germplasm collection, maintenance and use. Elsevier Press, Amsterdam.

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

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, 156: 847-854.

Chang HL, Fernando RL, Grossman M (1991). On the principle underlying the tabular method to compute coancestry. Theor. Appl. Genet. 81: 233-238.

Chatwachirawong P, Srinives P (1999). Coefficient of parentage of major sugarcane clones in Thailand SABRAO J. Breed. Genet. 31: 51-57.

Cho Y, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch S, Parl W, Ayers N, Cartinhour S (2000). Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. 100: 713-722.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics, 21: 3674-3676.

Cordeiro G, Casu R, McIntyre C, Manners J, Henry R (2001). Microsatellite markers from sugarcane (Saccharum spp.) ESTs cross transferable to erianthus and sorghum. Plant Sci. 160: 1115-1123.

D'Hont A, Rao PS, Feldmann P, Grivet L, Islam-Faridi N, Taylor P, Glaszmann JC (1995). Identification and characterisation of sugarcane intergeneric hybrids, *Saccharum officinarum* x *Erianthus arundinaceus*, with molecular markers and DNA in situ hybridisation. Theor. Appl. Genet. 91: 320-326.

Duarte FL, Silva P, Santos J, Barbosa G, Ramalho-Neto C, Soares L, Andrade J, Almeida C (2010).genetic similarity among genotypes of sugarcane estimated by SSRand coefficient of parentage. Sugar Tech. 12: 145-149.

Feng S, Li W, Huang H, Wang J, Wu Y (2009). Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). Mol. Breed. 23: 85-97.

Fraser LG, Harvey CF, Crowhurst RN, Silva HN (2004). EST-derived microsatellites from *Actinidia* species and their potential for mapping. Theor. Appl. Genet. 108: 1010-1016.

Gao LF, Jing RL, Huo NX, Li Y, Li XP, Zhou RH, Chang XP, Tang JF, Ma ZY, Jia JZ (2004). One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. Theor. Appl. Genet. 108: 1392-1400.

Garcia AAF, Kido EA, Meza AN, Souza HMB, Pinto LR, Pastina MM, Leite CS, Silva JAG, Ulian EC, Figueira A, Souza AP (2006). Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. Theor. Appl. Genet. 112: 298-314.

Gawel N, Jarret R (1991). A Modified CTAB DNA Extraction Procedure for *Musa* and *Ipomoea*. Plant. Mol. Biol. Rep. 9: 262-266.

Glynn NC, Mccorkle K, Comstock C (2009). Diversity among mainland USA sugarcane cultivars examined by SSR genotyping. J. Am Soc. Sugarcane Technol. 29:36-52

Grattapaglia D, Sederoff R (1994). Genetic linkage laps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: Mapping strategy and RAPD Markers. Genetics, 137: 1121-1137.

Gupta P, Rustgi S, Sharma S, Singh R, Kumar N, Balyan H (2003). Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol. Genet. Genomics, 270: 315-323.

Gupta S, Prasad M (2009). Development and characterization of genic SSR markers in *Medicago truncatula* and their transferability in leguminous and non-leguminous species. Genome, 52: 761-771.

Jaccard P (1908). Nouvelles reseachers surla distribution florale. Bull. Soc. Vaud.Sci.Nat. 44:223-270

Jackson PA (2005). Breeding for improved sugar content in sugarcane. Field Crops Res. 92: 277-290.

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

Kantety RV, La Rota M, 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.

Kawabe A, Miyashita NT (2003). Patterns of codon usage bias in three dicot and four monocot plant species. Gene. Genet. Syst. 78: 343-352.

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.

Maliepaard C, Alston FH, Van Arkel G, Brown LM, Chevreau E, Dunemann F, Evans KM, Gardiner S, Guilford P, Van Heusden AW, Janse J, Laurens F, Lynn JR, Manganaris AG,Den Nijs APM, Periam N, Rikkerink E, Roche P, Ryder C, Sansavini S, Schmidt H, Tartarini S, Verhaegh JJ, Vrielink-van Ginkel M, King GJ (1998). Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. Theor. Appl. Genet. 97: 60-73.

Ming R, Moore PH, Wu K-K, D'Hont A, Glaszmann JC, Tew TL, Mirkov TE, Da Silva J, Jifon J, Rai M, Schnell RJ, Brumbley SM, Lakshmanan P, Comstock JC, Paterson AH (2010). Sugarcane Improvement through Breeding and Biotechnology. John Wiley & Sons, Inc.

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

Mun JH, Kim DJ, Choi HK, Gish J, Debelle F, Mudge J, Denny R, Endre G, Saurat O, Dudez A-M, Kiss GB, Roe B, Young ND, Cook DR (2006). Distribution of microsatellites in the genome of *Medicago truncatula*: A resource of genetic markers that integrate genetic and physical maps. Genetics, 172: 2541-2555.

Oliveira K, Pinto L, Marconi T, Margarido G, Pastina M, Teixeira L, Figueira A, Ulian E, Garcia A, Souza A (2007). Functional integrated genetic linkage map based on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross. Mol. Breed. 20: 189-208.

Oliveira KM, Pinto LR, Marconi TG, Mollinari M, Ulian EC, Chabregas SM, Falco MC, Burnquist W, Garcia AAF, Souza AP (2009). Characterization of new polymorphic functional markers for sugarcane. Genome, 52: 191-209.

Pinto LR, Oliveira KM, Marconi T, Garcia AAF, Ulian EC, De Souza AP (2006). Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs. Plant Breed. 125: 378-384.

Pinto LR, Oliveira KM, Ulian EC, Garcia AAF, De Souza AP (2004). Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. Genome, 47: 795-804.

Price S (1963). Cytogenetics of modern sugar canes. Econ. Bot. 17: 97-105.

Ritter E, Gebhardt C, Salamini F (1990). Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. Genetics, 125: 645-654.

Rohlf FJ (1998). NTSYSpc: numerical taxonomy and multivariate analysis system, Exeter Software, Setauket, New York, 2: 02.

Rodriguez H, Castillo M, Flores B (2005). Genetic diversity of the most important sugarcane cultivars in mexico. e-Gnosis (Online) 3(1).

Rozen S, Skaletsky H (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. 132: 365-386.

Saha M, Mian M, Eujayl I, Zwonitzer J, Wang L, May D (2004). Tall fescue EST-SSR markers with transferability cross 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.

Selvi A, Nair NV, Balasundaram N, Mohapatra T (2003). Evaluation of maize microsatellite markers for genetic diversity analysis and fingerprinting in sugarcane. Genome 46:394-403.

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

Telles GP, Silva FRd (2001). Trimming and clustering sugarcane ESTs. Genet. Mol. Biol. 24: 17-23.

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

Varshney R, Graner A, Sorrells M (2005). Genic microsatellite markers in plants: features and applications. Trends Biotechnol. 23: 48-55.

Vettore AL, Silva FRd, Kemper EL, Arruda P (2001) The libraries that made SUCEST. Genet. Mol. Biol. 24: 1-7.

Wu KK, Burnquist W, Sorrells ME, Tew TL, Moore PH, Tanksley SD (1992). The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor. Appl. Genet. 83: 294-300.

Wu RL, Han YF, Hu JJ, Fang JJ, Li L, Li ML, Zeng ZB (2000). An integrated genetic map of *Populus deltoides* based on amplified fragment length polymorphisms. Theor. Appl. Genet. 100: 1249-1256.

Yan Q, Zhang Y, Li H, Wei C, Niu L, Guan S, Li S, Du L (2008). Identification of microsatellites in cattle unigenes. J. Genet. Genomics, 35: 261-266.

Supplemental Table 1. Functional annotation of 212 EST-SSR sequences.

S/N	Marker	Cluster	Motif	Forward primer	Reverse primer	Time (°C)	Expected size (bp.)
1	SEM1	SCEPRZ1008B02.g	(agc)10	ACCATCGCAATCGATGTTTA	AACTGGATGGCGTACAATCA	56	215
2	SEM2	SCACAD1035F09.g	(cg)7(ct)14	ACACCGAGCTGTCCCAAT	GCATCTGATGAGCCTGTGAA	56	164
3	SEM7	SCAGLB2047G03.g	(tttc)15	GGAAGGTATGGGTGCTATGC	ACAGGGCAATAACAGGGGTA	58	167
4	SEM11	SCAGRT2041A11.g	(tg)16	AGGGCTTGGAAGAAAGGAAT	TGGCAAGCAACAGCTAAAAC	58	190
5	SEM13	SCBFAM2022D09.g	(gaaa)11	GCGAGAGAAGCTAGGAAGCA	CGCAGATCCTCTTGAACCTC	62	161
6	SEM14	SCBGAM1091A05.g	(ta)10(cag)5	GAGCAACGAGCTGAAAAGTG	TGTCCTGACCTAGGATGTGC	59	245
7	SEM15	SCBGLR1119D12.g	(ta)26(at)5	TGTCCACAATTTTGGCTGAT	GTTGCTTGCCTGATCATTGT	56	200
8	SEM16	SCCCAD1003H03.g	(aag)15	CCTTCCTTGGCCTCTTCTCT	TGCTGGTCGCAGTACTTGAT	58	248
9	SEM19	SCCCCL4002A08.g	(gca)5(caa)13	CAGCCCATTAACCAAGCAAT	GAAGCAGCTGTTGCTCACTG	58	184
10	SEM20	SCCCCL4004D08.g	(ataga)12	CCGGCTGTGAAAATTAGGTT	TCGAATTGGTCAAGACTCTCC	58	219
11	SEM24	SCCCCL4015G07.g	(ct)10(ac)5	CAATTCGTGGCTTGTGTTTG	AGCAGAATCGGCAAGGTAAA	58	197
12	SEM27	SCCCCL6001F08.g	(ag)16	AGGTCTCGGCACTTGGAGTA	GATCGATCCTCCCTCTTTCC	62	172
13	SEM29	SCCCCL6005F10.g	(ct)21(ccg)7	GATCGATCCTCCCTCTTTCC	AGGTCTCGGCACTTGGAGTA	62	187
14	SEM36	SCCCLR1078G05.g	(ta)12	CCATGTGCAGCATTTAACAA	TGGACATGCTAATGACTACTGC	56	152
15	SEM37	SCCCLR1079D03.g	(ga)15	CTTCCTGCTTCGAACATTTG	ACGAGGTAGATCCCGAAGG	56	214
16	SEM38	SCCCLR1080G12.g	(ca)10	TGAATTGCACAAAACACCAA	GACGGTGTAAACAAGCTGTGA	55	249
17	SEM41	SCCCRT2001B06.g	(ag)24	CCCCCTTGACACCTCTGTAT	TAGAACGAACCAGACGACCA	59	248
18	SEM42	SCCCRZ2004G11.g	(aag)7(ta)15	GACTTCGGGAAGAAGGAGGT	ACCAAGCACATCCAGCAGTA	59	285
19	SEM46	SCCCST1002H09.g	(gct)11	GCGGTCTTGTAGTGGTAGAGC	TGATCCTCCTGTTCCTCACA	59	232
20	SEM53	SCEPAM1020C01.g	(ct)20(ct)20	CCGCCTTCTCCTTAGTGACA	CCACAAGCCTAATACAGCTCAA	62	192
21	SEM55	SCEPAM2057B09.g	(ct)15(ct)5	ACGGCATCAGATTCAGATCA	ATGGCTTTCCATCTCGTGAC	56	277
22	SEM57	SCEPRT2048F09.g	(cag)11	TCCAGAAGTACGTGGAGACG	ACGACAGCAGGTCGAACAT	56	222
23	SEM58	SCEPSB1130G10.g	(cacta)15	CCAACCAACCTCGACATTCT	CCATGTGATCTGACCTGGTG	60	209
24	SEM60	SCEQAM1041B03.g	(ag)12	TGCTAACACATTTCAAGAAAGAGA	GATCCAATCCGAGGAAAAGT	58	155
25	SEM61	SCEQAM1041H02.g	(ct)12	GTTCAGAACACGTGCAGCAT	CACGCTTGACATGAGAGGAA	59	165
26	SEM63	SCEQHR1078A09.g	(ag)25	GGTCGGTGCTCTGTTCTTTT	CCTGCAGCAGAGACGAGAT	59	197
27	SEM68	SCEQRT1032H07.g	(ta)6(ac)11(tc)6	CCCTGAGGTCTCTCTCCACT	TGCCATAGGACAAGAGTTTAACA	64	239
28	SEM72	SCEQSD2075F10.g	(ag)14	GAACCTTATCGGTAGCCTCCT	GAGCGCCATAGGAGAAGTG	58	158
29	SEM76	SCEZHR1048C09.g	(gata)14	ATATGCCATGCACGTGTGAC	CAATGTAACCATCGCAGCAT	58	190
30	SEM78	SCEZRT2023F09.g	(ggt)5(atag)14	GTGGTCGCAGACGAGGTC	CTCCGCATTAGCCATTTCC	58	222
31	SEM79	SCEZRT2024C04.g	(ac)13	GATGGAACAGATGCGACAAG	GTTCATCGTAACCTGCTGGA	58	207
32	SEM80	SCEZRZ3015G05.g	(ta)27	GCAGATGAGAGGGCAAAAGT	CGCCTGCAGATGAATCATAG	59	244
33	SEM82	SCEZRZ3017G04.g	(ga)12	AGTACAAGGCACAGCCAGAG	GGACATGAGGTACACCCAGA	62	197

Supplemental Table 1. Continue.

34	SEM83	SCEZRZ3096G10.g	(atg)10	TCCTCCTCTTGTTGCAGTTG	GTCGTCGTCACGATCATCTC	58	184
35	SEM84	SCEZSB1094A08.g	(tc)10	TGTAGCAATTCCTTGCGTTG	CAACAAATACAATGCCAATCG	58	229
36	SEM85	SCJFAD1011F07.b	(ga)10(agag)5	CACCTAGTGAAAGGGGCAAA	CCTGAAGCCTTGGTAGCATC	59	242
37	SEM86	SCJFLR1035E04.g	(ga)10	CGAGAACTAGCATAGCACAAGA	AACAACTGGTGCAAGTCCAT	58	172
38	SEM90	SCJFRZ2010C01.g	(ag)17	CGTGGGCTAATACAGCTCAA	CCCACTGGCACAATCTTCTT	60	190
39	SEM92	SCJFSB1010B12.g	(tc)10	TGTGCCGTTGCCTAATAACA	GGCAAGCTTCCTCAGTTTCTT	56	209
40	SEM94	SCJFST1014E07.g	(tc)10	TTATCTCTTCCCTGCGGTCT	CTGGCTGGGGAGCAACTT	56	196
41	SEM97	SCJLLR1033A04.g	(cgt)10	TTTGGGGGTTCTTAGTCCAG	GAAGACAGTGGACGAGGTCA	59	233
42	SEM98	SCJLRT1019D02.g	(tg)15	GCCAGCAGAATGCTTAACAA	TCTTACTCTGTCCCCCAACC	58	233
43	SEM99	SCJLRT3076A02.g	(ta)12	TTACCTCCGGCAACGTTAAA	TGCAGGCATATGGTAGTCCA	56	175
44	SEM105	SCMCAM1100G01.g	(ctttt)18	AGGGGCCTCAAGTTGTTCTT	CGGTCTCATGGTCACCTTTT	58	220
45	SEM106	SCMCCL6027C07.g	(tc)24	AGGTTGCTGATGGTCCTCAC	CAAGAAGGGAAGCAGGGACT	62	197
46	SEM108	SCMCFL5008F03.g	(ag)10	CGACTTGTGTGGAGGTGAAA	TGGATCAATGTGAACAAAATCG	58	164
47	SEM110	SCMCRT2103A12.g	(at)12(tg)25(ccg)5	GTCTCCTCCTCAACCACTGC	GTGGGGTATGTAACCCATGC	62	202
48	SEM112	SCMCST1057D03.g	(tc)14	CCTTCTGCAGACGAGTTGAA	ACCTGACCAGCAAATCAACA	58	159
49	SEM113	SCPIRT3024F01.g	(ga)17(gga)5	TTCCGGTTTACCCTGCATAG	TTCCTCAGGGCCCTTTTATT	58	232
50	SEM117	SCQGLV1018G08.g	(at)41	GCGTGGCACTGACTACAAGT	CAATGTTCTGTGGTCTGCACT	56	246
51	SEM119	SCQSHR1020F04.g	(ct)16	CCGCGTGCTCTCTCTCTCT	ATTGCCATCACCTCATGCTT	56	206
52	SEM120	SCQSRT2032H08.g	(ccg)10	TAGCAGCTCGATTCACGATT	ATCCTAGTGTGGTGGGTGCT	58	176
53	SEM124	SCRFLB1055F01.g	(tcc)11	CCTTGATGTGCTTGACGAG	CCAACGAGCAAAAGTAAACG	56	270
54	SEM125	SCRFRT3058D07.g	(ct)16	TATTCTCTCCGGATCCCCTA	ATTCAAAGCGCAACACAGTC	58	160
55	SEM127	SCRLAD1139G03.g	(ag)23	GCAGGATAAGCTTGGCAGTT	CTCAAGGCAAGGAACGAATC	58	286
56	SEM131	SCRUFL4024B04.g	(ac)18(cgc)6	AAAGGAAAGCAAACCCAAGG	GCAGTCGTTGTCGTAGCAGA	58	245
57	SEM132	SCRURT2010E12.g	(ct)18	CCCTCCACCTCTTTGCTC	TAGAAAGACCTGCCCTCCTG	58	202
58	SEM136	SCSBAM1085C09.g	(cgg)12	CTCTGACCCGAGCAAAGG	GCCAATAAACAGCAGGGGTA	56	238
59	SEM140	SCSFAD1114H02.g	(at)23	GTGTTTTGGAGACCGTGTCA	CGATTGTTGCGCTGTACATC	58	127
60	SEM141	SCSFFL4083B01.g	(ac)16(ag)13(gt)5	ATCATCCACAGCTAGCAGCA	GGTTTTGCCTTGGTTTTTGA	56	174
61	SEM143	SCSFHR1045G08.g	(ct)9(ct)12	CCCCTCTCCCTCAGTCTTCT	CCATGCTGTCAGGATCCAC	60	279
62	SEM147	SCSGHR1070F11.g	(tcc)10	GCCTCTTCCTCCTCCACTC	GACGACCGTCCTTGTTGAG	58	250
63	SEM148	SCSGLR1025D03.g	(tca)12	GCTACCGGATGGATAAAAGC	CTGACCGAAATGATCAAGGA	56	245
64	SEM154	SCUTCL6035D02.g	(ct)14	ACCGAGGTAGGAGGGAGTGT	GCTCGCCATGAATAGAAAGG	58	219
65	SEM158	SCVPCL6044A06.g	(ag)12	GGATGGTTAAAGCGGAAACA	GGAAACAGTGTACGCCCAGT	56	151
66	SEM159	SCVPHR1089A09.g	(ag)13	CTGGTGGAATAACTCGCTGA	CTCAAGGCAAGAACGAATCC	60	279
67	SEM161	SCVPHR1094C01.g	(agg)10	GAACTGCTCACTGGCTCCTC	GTAGAAGTCCGTCGCCGTAA	58	186
68	SEM164	SCVPLR2019H04.g	(ag)18	GAGGTTATGGGGAAACCAGA	GATTGCAGCCGTAAACTTGA	56	211
69	SEM166	SCVPRT2080G09.g	(gt)22	AGCGCATCTTGCTTATTTGA	ATGCATGATCATCGAGGAAG	56	171

Supplemental Table 1. Continue.

70	SEM167	SCVPRT2083D03.g	(gt)30	CGGATCTTGGCTCCTTCTCT	AGCCTTGATTGGCAATGGTA	58	223
71	SEM168	SCVPRZ2037E02.g	(ct)11	AACGTCAGCCGCTACAACTT	CTTCCCTTTTGCGAAGAAAA	56	143
72	SEM174	SCQGFL3059G12.g	(ta)11	CTCACCGCAGCTCTTTTTCT	CACAAGCTATGCGGTCAAAA	58	249
73	SEM176	SCRFFL1029H06.g	(cctc)5(ct)10	CGCCATAACCATAACCACAG	CCTCCCTCCGCTACTTCCTA	58	188
74	SEM179	SCSBFL1101G01.g	(aaag)10	TATTCCACCGGGAACAAGAA	GGGATTGTAGCGACGAGTTG	58	210
75	SEM180	SCSBFL1104E01.b	(ct)12(tgc)6	TTCCACATCAAGCAAGCAAG	ATGACATCAGGAGGGAGACC	58	206
76	SEM184	SCVPFL1073A11.g	(ga)10(ggc)7	ACCAACGCGACGAGAGAG	GCCTGAACTGGTCGTAGGTC	58	206
77	SEM189	SCJFAD1013E12.g	(agg)6(ggaa)8	GAACTGCTCACTGGCTCCTC	GTAGAAGTCCGTCGCCGTAA	59	209
78	SEM190	SCRLAD1098A04.g	(ct)9	CTAGCACGGCAATACAGGC	AGATCTGTTGGGTGCTCGTC	59	189
79	SEM191	SCRLAD1138A05.g	(gcc)8(cca)5	CCAGTCGCGATTCTTCCAC	AAGGGACGGGGAGAAAAATA	55	176
80	SEM195	SCSGAD1008F08.g	(ggc)9(gag)5	CTTCCCGTCGCTCTTACCT	CTCCTCCTCCTCCTCTCCAC	60	183
81	SEM199	SCEPAM1021B02.g	(tg)9	CTCTCGAGGAGGTGGATGAG	CTGCAAGTTTGTTGGCTGAA	56	237
82	SEM200	SCEPAM1050A03.g	(tc)9	CTGCAGGATCACCTGGAAC	TAAACCCACGCTGACAGACA	56	238
83	SEM203	SCEQAM1036D03.g	(cgc)6(cgc)8	GCGGCCTCATACGTGTAGAT	TCTCTTCCCCTCACCAGAAA	56	237
84	SEM206	SCVPAM1056A04.g	(tctt)9	CATGGTAGCTCCGCTTCTTC	GCGAGAAGCTAGGAAGCACA	59	197
85	SEM207	SCACAM2043G03.g	(gct)8	GGCACACCTCGAGAGACC	ACTCCTCCTCCTCGCTTAGG	60	151
86	SEM208	SCACAM2044B11.g	(ca)8(ac)5	TGAAGACGATGATGGGATGA	TCTGTTTTGCTCCTCCGTCT	56	302
87	SEM211	SCCCAM2001E04.g	(cgc)8	CGGTCGTCTCTTCCTCCTC	CTACTACCACCCGGACCAGA	59	212
88	SEM213	SCCCAM2C08B11.g	(tc)5(tc)5(ct)5(ct)6	CTCTCCGACTCGTCTTCCAC	GCGGACTGCAAAAGAGAGAT	56	241
89	SEM214	SCEQAM2037C11.g	(cag)8(cg)5	ATCGGCTCCAGTCAGAGAGA	CCTGGTGAAGGCTCATGATT	59	323
90	SEM215	SCEZAM2033H10.g	(tc)8	GCCGAAGAGGAATCTACGAG	GTTTGTCTTCCTCCCTGTGC	56	193
91	SEM217	SCEZAM2096F07.g	(ga)9	CACGGGGAGACGAGAGAC	CCAACAACAACCAGAATATCG	54	174
92	SEM219	SCMCAM2084A04.g	(cag)6(cag)5(gca)5(cag)5	AAGTACGGAGCGCAGTGTAG	ACCGCCTTGTACTCCAAATC	56	228
93	SEM220	SCMCAM2084F10.b	(at)8	AAGCTCCTTGCCTGCTACTC	CAAAGGGCATCCTTTCTGAT	55	218
94	SEM221	SCQGAM2028B01.g	(ccg)8	GCCTCTCTCTGCTCAGCCTA	CTCCTCATCTCTCGCCAAA	56	170
95	SEM223	SCSGAM2076E10.g	(tc)8	CACAGCACTTGCCAAGCTAA	AGTTTCACAAAGGGCGACTG	56	216
96	SEM227	SCCCCL3005D03.b	(ct)8	GCTACAGTGCCTCTCCCTCT	CTAGAAGCAGAAGTGGAGTGCT	59	287
97	SEM229	SCCCCL4007E05.g	(cga)9	AGAACCACAACCACCAGGAG	ACAGTTGAATAGGCCGGATG	58	343
98	SEM231	SCCCCL4013B10.g	(cgt)5(gcg)9(cg)5	CCGTTCTACACCTCCAACAT	GACCGTGACCATCTGCTG	57	426
99	SEM232	SCCCCL4014F09.g	(ga)8	CAACTCCAGCTCCAGTCTCC	CTTTTCGCGAAGTGAACACA	58	311
100	SEM233	SCCCCL4015B01.g	(tgt)8	TTGCTTGGGACAAAAGGCTA	ATCTTGCAAAGGAAGGAGCA	55	336
101	SEM234	SCACCL6009D08.g	(acc)8	GGACATGCTGCTCCCTACAT	AGGAGGACTGGTGGTTGAGG	60	211
102	SEM235	SCACCL6010C05.g	(tc)9	CATCGGCTCATCATAACGAA	AGCTACTTCAGCCCCAAGTG	55	250
103	SEM236	SCCCCL6003H04.g	(ct)9	CCCTTTGCTTCCCCTTTACT	GAGGCGCCTTACTGTTCTTG	56	193
104	SEM237	SCEPCL6023G01.g	(gca)8	AGGGAAAGAGACGAGGGAGA	CGTATCTCCGACCACTCCAC	59	167
105	SEM238	SCEPCL6029D06.g	(gcg)8	CTCTCCCCCAACTCTCTCTG	TCCGACGTCAACGTCTCAG	59	176

Supplemental Table 1. Continue.

106	SEM239	SCRLCL6030D09.g	(ct)9	CGAGAAACCGTGTCCCCTA	CCCTCTCCCTCTTCCTCCT	59	155
107	SEM246	SCJLFL4097F08.b	(cg)5(gaaa)8	AATCGATCTTAGGGCCGGTA	ACGCCGACGAGTGAGGAC	58	276
108	SEM250	SCRLFL4109G12.g	(ga)8	ACGACTGTTTGTTCGTGCTC	TTCAAAGGGGCTATCTTGCT	55	233
109	SEM254	SCBGFL5080G03.g	(cgg)8	ACCTTACAGAGCCCACTGCT	TCGCGATAATGAGATTGAGC	59	152
110	SEM255	SCCCFL5062D10.g	(ag)8	CGGCGTCCACTGAAAGAG	CAGCCTCGAGTTGGGATG	56	178
111	SEM257	SCEZFL5084A01.g	(tg)8	TGCTGGAGACGGAGTAGCTT	ATCAGGCAAGCACACAATCA	57	159
112	SEM258	SCEZFL5091D04.g	(cac)9	GGAAGAGGAGGCTTCGAGAT	CTGGATAATCACGCCCAAAT	55	343
113	SEM261	SCAGFL8042E05.g	(cgg)8	CCATCCATCCTCTCATCTCC	AAGAGTGCTTGAGCGGATCT	56	187
114	SEM263	SCRLFL8053B05.g	(gga)5(gca)8	AGCCTCTGACGCTAAGATCC	CACACGCTGCAGATGTTGTT	56	208
115	SEM265	SCAGHR1018C11.g	(cca)8	ACACTAGCTAGCCAGCCACA	GAAGCGAGGCTATGGCTATG	57	163
116	SEM271	SCJFHR1034E09.g	(ccg)9	AGCAGATTCACTTCGCCACT	CGATGAGCTTGGAGAGGAG	55	157
117	SEM273	SCQSHR1022B03.g	(cat)9	TTTCTTTTCGTCACACCCAAT	ACTCCCGTCACTCACCTGAC	55	180
118	SEM275	SCRUHR1074E09.g	(gag)9	TCTCATCGGATTCACACACA	GGGCAGCTTCGTAATGGT	55	242
119	SEM276	SCSFHR1043F12.g	(tg)5(tg)8(ag)5	AACCCGTTCTTCTTCCCCTA	CAGAGGGAGATTTGCCATGT	55	241
120	SEM279	SCVPHR1092G06.g	(gt)8	AACCTAAACGACGACGATGG	AGCGAGGAAACGTCGTACAT	58	343
121	SEM280	SCCCLB1002D05.g	(ac)5(tac)8(ct)7(act)5	ATGGAGCTCCGTCTTCTTGT	AGTACCGTAGTTGGGGGTTG	54	222
122	SEM282	SCQGLB1038F11.g	(gcc)9	GAACCTCGCAGTCTTCACAA	CACTACCTGCCTTTCTCTCG	56	191
123	SEM285	SCVPLB1020B05.g	(gcc)9	TCCTTGAACCTCGCAGTCTT	CTACCTGCCTCTCTCGTTCC	54	192
124	SEM287	SCRULB2062C01.b	(ac)5(ac)9	CCAATTCAACAAGCATCGAG	GGGAGGACATGAAGTCTGGT	59	187
125	SEM288	SCACLR1057E07.g	(ga)8(gct)5(ctg)7	TCCGATCACAATCACAGACC	GCTGCAGCAGATGACAAACT	58	227
126	SEM290	SCBGLR1002F11.g	(cgc)9	AAACGCAAACCCTTATCTCG	GCTTGGAGGTCACCTTCTTC	55	250
127	SEM294	SCCCLR1066D07.g	(cca)5(ccg)8	CCATACCCTGTACCGTACCC	GATGCTTGCATTCATCCTTG	55	184
128	SEM295	SCCCLR1066F12.g	(tc)8	CACCTCCCAGACTCTTCTCC	GTGACACCATGGTCCTGAAG	56	176
129	SEM297	SCCCLR1075D10.g	(cgc)8(cac)6	CACCAAACAGACTCGCATTT	CGGATCGAACTCTGTGACAT	54	214
130	SEM298	SCCCLR1076A04.g	(ca)8	ACGCGAGAGGGAGAGAGATA	GTCAGCAGCACGAACAGC	59	163
131	SEM302	SCEPLR1008H10.g	(ggc)5(cgg)8	GCGGTTTCTTGTTTTCCTTC	ACCACGACCTCGATCTCAAC	56	282
132	SEM303	SCEPLR1030D11.g	(agg)8	CGAAAACCCTCAAACCCTAA	CTCCTCTAGCTTCCGCTTGT	55	193
133	SEM306	SCJFLR1013A08.g	(cag)8(cag)9	ACCACCAATACCACCACCAC	TCGACGTTGGACTTGAGAAG	56	247
134	SEM307	SCJFLR1074A10.g	(ac)5(ca)9(ag)7	CAAACTTTTGCCCGATAGGT	CGGAGCATACCAAGTGAAGA	54	244
135	SEM308	SCJLLR1101F02.g	(gt)8	TCTCGACTCCCCTAATCACC	CGGACAGAAAGATCGCAGTA	55	242
136	SEM310	SCQGLR1019C10.g	(ga)8	AAGAAACCAACCCTCAAAGC	GTAGGGTAGCGCTGGGTAAT	55	230
137	SEM313	SCSGLR1084A02.g	(gcg)9	GAGGGAACACATCCCTTCTC	GCCGTAGATGAAGACCTCCT	56	211
138	SEM314	SCVPLR1049G12.g	(ct)8(ct)5(cgc)7	GAATATAACCGCCACCTTGC	TGGCTTTCCATCTCGTGACT	58	250
139	SEM315	SCACLR2007A01.g	(ca)5(ag)9	GAGGTCCTGGGAGAGACAGA	GTCTGGCCCGTAAGCTGT	60	152
140	SEM319	SCAGLR2026C05.g	(cgc)8	ATCGTCATCGCAAAATGC	CAACCGGAGGCACTGAGTA	56	150
141	SEM320	SCCCLR2002F05.g	(ag)9(ag)7	GAGGCAGCTCGACGACAC	GTCAGCTCCGCTCCTGCT	62	111

Supplemental Table 1. Continue.

142	SEM321	SCQGLR2032D06.g	(ga)8(gc)5	GTCCGTCTCCACTCGAAAAC	GCGGTTGAGGTCGAGGTAG	59	222
143	SEM327	SCCCRT1001G10.g	(cct)8	CTCCCCTCTCGCTCATCA	AGGTTGACGATGGTGGTGAC	59	196
144	SEM328	SCCCRT1003H03.g	(ct)16	TCTTGCCTGTTCGTCTTCCT	ATTCCGATTCCGATTCCAAC	55	239
145	SEM329	SCEQRT1025C10.g	(cgg)8	CACCCAGCTCAAGTACAGCA	GCCTGTAAAAGCCTCCTGTG	59	222
146	SEM332	SCJFRT1009B09.g	(ggc)8	CCGCAAGGAAGAACACCTT	GCAGTGGAAGTCGACGTAGG	56	232
147	SEM336	SCJLRT1006C08.g	(at)5(aag)8	GCCAGGGTTCTTCAAGTGAT	TTCGTCATAGCCATCGTCAT	55	155
148	SEM337	SCJLRT1013F12.g	(ga)8	AGCAATGGTACGCACAAGAG	TTGCTAGTCGTCGTTCTTGG	55	202
149	SEM338	SCJLRT1018G02.g	(ga)8	GATCGGATCGAGAGGAGTTTT	ATACGACGAGGACGAAGTGG	56	216
150	SEM339	SCJLRT1019C06.g	(ag)8	AAGCGAGCGTACACCAAATC	ACGGCTCAGATGGTTGAGAG	58	163
151	SEM341	SCAGRT2041D09.g	(cgg)8	GTGGTTTGAGTACGCTCGTG	AGAGGGATGGCAGTATCCAG	56	249
152	SEM344	SCEPRT2047A05.g	(ct)15	CGTGCGCTCTCTCTCTCTCT	ATTTTGAGATGGCTGCATCA	57	171
153	SEM349	SCEQRT2099E08.g	(gca)8	CGAGAGGCCTTCTCTCTCTG	CGCTGACGTAGTCCTGGTAG	59	180
154	SEM350	SCJFRT2057F04.g	(gca)9	CCAATGGAGACGACACTCCT	GCGGACGTAGATGGAGAAGA	59	228
155	SEM351	SCMCRT2085E08.g	(tg)8	CGACTGTGGGAGGAGTTTGT	TTGCAGCAGTTGCTAGCTGT	57	119
156	SEM353	SCQSRT2033C08.g	(ttg)9	TTGCTTCTGTTGGGTTTCAA	TGGTTAAGGTTTGTCGGTGA	54	191
157	SEM355	SCSFRT2067E08.g	(ag)8	ACCAAATCCAAACACAGCAG	CGATGGTTGAGAGCTTGTGT	57	144
158	SEM358	SCAGRT3048C12.g	(gaa)9	CTGGCCTCAAGAGGAAACTG	ACCAACCTCTTGACCAGCAC	59	124
159	SEM361	SCCCRT3001D09.g	(ct)9	GTAGCCGTGGAGCATGAAGT	CTGCTGCCATTAGGAGCAAT	59	173
160	SEM366	SCBFRZ2045C02.g	(ca)9	CCACCTCTTCTGCCAAGAAC	CATCTTAAACTCCGGTCCACA	55	167
161	SEM367	SCCCRZ2001C02.g	(ag)8	AGTCAGCATCCATCCAGTCC	ATTTCTCCTGCCCTCCTCTC	59	196
162	SEM368	SCCCRZ2004C05.g	(gcg)9	AAACCCTCGCCTCCGATT	CCCAATGGTACCAGCAGAGT	59	241
163	SEM369	SCJFRZ2015A10.g	(ga)8	CGCTTCCATATCTTCTTCTTGG	TGACTCTCCGGTCCCTACAC	55	123
164	SEM371	SCJFRZ2034B06.g	(tg)9	GGAGAAGCATTTCAGCAACC	CCCGCTTTTCCTCTTTCTTT	54	238
165	SEM372	SCVPRZ2036E01.g	(at)8	GCCAAGCTAAATAGCTGCTG	ACCACCGTTTCTTTCCTGAC	57	205
166	SEM373	SCVPRZ2038E05.g	(ccg)8	GCGACCAAATCTGCCGTAT	CATGTAGTCGAGCGCAGAGA	58	189
167	SEM374	SCCCRZ3003B01.g	(ag)9	GCCTCCTCCCTCCTCTTCTA	GACTGGCTCGGAAACCCTA	59	141
168	SEM375	SCEPRZ3128D05.g	(ct)6(tc)9	ATGGAGGCTCGTTGTCTTTG	CCGTAATCGCCTCCACTAAA	55	174
169	SEM377	SCEQRZ3020E12.g	(gcg)8	GGAGAGGACGAAACCCTAGC	CGCATTGAACGCAGTTTCTA	55	233
170	SEM379	SCJFRZ3C03A08.b	(ctgtg)9	ACGAGGCCACCATAGAACAT	GCACAAGGTGATTGTGCTGT	56	221
171	SEM384	SCUTRZ3103F01.g	(cgg)9	TAGTAGCAAGCGAGGCGATA	GTCTGTTGCCTTTGATCGTG	55	228
172	SEM390	SCSGSB1005B12.g	(ag)9	GGGGAAGTAAGTCTCAGGTCA	GCCACCACCTCCATTATCTT	57	116
173	SEM391	SCUTSB1033C02.g	(ag)8	GTTCAGACTCGCGTGTTTTT	GCTGAGAACCCTTCAGCTCT	55	104
174	SEM392	SCUTSB1075H09.g	(ta)8	TCATGCTCACCAGCAAAGAC	TCCCGATCAGTGTGTAGACG	55	234
175	SEM398	SCEPSD1006D03.g	(ta)9	CGTGCAAGCTCCAATATGAT	TGCCACTGTATAGCAGCGTA	54	184
176	SEM400	SCEZSD1081A02.g	(ccg)8	CAGCTCATCCTCGTCAACCT	CTCCTCTGCTCCTTGTTGCT	59	225
177	SEM401	SCMCSD1059G09.g	(ct)9	GCTCCATTCATTTCCTCCTC	TTCGATCGATTGATGGTTGA	53	111

Supplemental Table 1. Continue.

178	SEM403	SCEQSD2077B12.g	(cga)8	CCTGCATCAACCTCTCCAC	GAAGGCGAGAGAGAAGATCG	55	242
179	SEM407	SCCCST1006B01.g	(cga)8	GCGAAACTAGCGCTGCTAAA	GGAGGTTCGGGTACGAGTC	58	295
180	SEM408	SCJFST1048G04.g	(ga)9	CAGAGCCAGCCAGGTAAAAG	TCATCGTGTGCTGCTGGT	58	228
181	SEM412	SCJLST1022C09.g	(ct)8(ga)7	CAAGGCTGCTTCTGGTGTC	CCTCTTTGGGTTCTCTGCTC	58	246
182	SEM415	SCMCST1050H06.g	(tc)8	CAGCAGACGAGACGAGAGAG	AGGGTGATGAAGGGAATGAG	56	163
183	SEM417	SCQGST1032E05.g	(agga)9	GTCTCCTCCCCCTCCTCTC	AGAAGGAGTCGCTCATCTCG	60	167
184	SEM418	SCQGST1032G11.g	(gt)8	CGGACGTCTCATGTTCTTTG	CAGTGTCCAGTGCAAGTTCC	55	244
185	SEM419	SCSFST1066E06.g	(gcc)8	TGCGTGGTTGATTGAAGAAG	AGAAGCCTCTTCTGCTGCTG	60	199
186	SEM421	SCSGST1069F04.g	(gga)5(ctc)8	CACCCTGCTGGTCTCCTC	TCGACGTCGTGTAGTGAACC	59	170
187	SEM422	SCSGST1072B03.g	(ag)8	GAAGAGTGGGGACGTCTCAG	GCCAGAGGATGTGGTAGAGG	59	199
188	SEM425	SCSFAD1070E12.g	(gcc)5	GTGCCACCAGCAGCAAT	TCTCGTAGCTGCTCGACTTC	56	244
189	SEM426	SCVPAM1059C01.g	(at)5	TCGAGAGCGGTTTCATCTTT	CTTTCCTGTCAGCCAAGTGA	56	471
190	SEM427	SCSBFL1105H11.b	(ca)6	AAGTAGCGGAAGCATTAGTTCA	CCAAGTTCCTCCTCACCAGTA	57	277
191	SEM428	SCSGLV1008C05.g	(cg)7	CAGGAAGAAACAGTAGGAAGCA	AGGTACTTGGCGGTCTTGAT	57	178
192	SEM430	SCRUSB1064F09.g	(cgg)5	TCCGACTACCTCAAGTGCAAG	GACGGCATCTTCTTCTTCTCC	55	224
193	SEM432	SCJLST1019B07.g	(gc)6	CGCGTCCGTAGATTAGTAGCTC	AGCGAGTAGATGTTGATGACCC	56	195
194	SEM433	SCSBST3094H07.g	(cga)6	GACACGCCCAAAGGAAAAG	GAGATCCGGACACACATGG	54	245
195	SEM434	SCEZLB1007E12.g	(ta)7	TTCTTGCTTCTTTCTTTCCGTC	TCAAATCGTGCTTGCTTGAG	52	236
196	SEM435	SCQGLR1041A05.g	(ga)5	AGGCTGAGAGAGCAAAGAAAGA	CCTAGGATCCTTCGGGTTTC	55	164
197	SEM436	SCJLRT1021D04.g	(tcc)5	GGTCCCATACATAACACAAGCA	TGCATGAAGAAGCTCAGGTG	57	248
198	SEM437	SCQSRT2031C10.g	(tc)5	CCTGGTTCCTGCACTTGTCT	CATCACTTGCCATCTGCATT	57	217
199	SEM439	SCACSB1117C07.g	(cgg)6	CGTCAAGCTGTAGTCCGAGAG	CTCGTCCCAGACCAGGAG	59	197
200	SEM440	SCACSD1018E05.g	(gac)5	AGCAACCTAATCACAGCAACAA	CCATCATCCGATCATCCTTC	56	229
201	SEM442	SCMCST1057C10.g	(gct)5	CATTTATTTGCCACCTAGAAGGG	AAACAGAAACCGGACAGCAC	56	195
202	SEM443	SCRLAD1043B06.g	(ggt)7	GGAATGGGAACAGCCACTAAC	AAGAAGGCTATCGAGGTGGG	55	323
203	SEM444	SCBGAD1027C03.g	(ggc)7	CACGGTTCTCCTGCTGAAAG	GACGGGGTTGTTGAAGGTG	55	313
204	SEM446	SCCCCL3001D10.b	(ccg)5	GAGCAGTCCCTTGCCATGT	GCCGTCGAGTACACCGTC	59	389
205	SEM447	SCEZFL5083C02.g	(gc)5	TGAGTTCAGTTCCTTCCCC	AGAACTCCAAGGAGCAGCAG	56	300
206	SEM449	SCEZLB1006B07.g	(gcc)5	TGGTGTGAGTTAGTGCCTGAGT	TAGAAGGTGTTGATGATGAGCG	55	265
207	SEM450	SCEZLB1007E12.g	(ta)7	TTCTTGCTTCTTTCTTTCCGTC	AGATGAACACATAGTTGCACCG	56	189
208	SEM453	SCBFRZ2045E11.g	(ggc)5	AGCGACATGAGCTACCGTCT	TAGTACCGCGACAGACCTTTCT	58	287
209	SEM454	SCSBRZ3122D09.g	(gga)6	GTAACTAGCAGCAACCCTAGCC	ATCCTCTTTTGCCTCCCCT	55	387
210	SEM456	AY302083	(tgc)6	TCGTCCTACAACCACGACTACA	GAGAGGCAAGCAAGGAAAGAT	56	164
211	SEST3	SCSFSB1097B02.g	(ta)8	CCCCGAAGATCAAGGATAGG	CGCATCTCAAATGGGAAAAT	56	413
212	SEST4	SCRLAD1040D08.g	(at)5	CAGGCACTGATGTCATGGAT	GAACTACACTCGCCGCTCAC	56	313

Supplimentary Table 1. Contd.

S/N	Marker	BLAST	E-value	PIC	Allele	Marker number		Segregation ratio	
					number	S6	Phil6607	1:1	3:1
1	SEM1	hypothetical protein	3.0E-34	0.64	9	1	6	6	1
2	SEM2	glycosyl hydrolase family protein 17	6.0E-19	0.9	7	1	3	3	1
3	SEM7	Hit not found		0.61	8	1	5	5	0
4	SEM11	hypothetical protein OsJ_018777	1.0E-11	0.93	11	2	7	6	1
5	SEM13	lysine decarboxylase-like protein	3.0E-50	0.68	7	2	2	4	0
6	SEM14	Hit not found		0.83	10	4	3	5	1
7	SEM15	Hit not found		0.45	5	1	1	2	0
8	SEM16	harpin-induced protein 1 family (HIN1)-like	5.0E-14	0.91	3	1	0	0	0
9	SEM19	ARFE_ORYSJAuxin response factor 5	1.0E-177	0.86	13	2	7	7	1
10	SEM20	NADH-plastoquinone oxidoreductase subunit K	1.0E-100	0	6	2	3	4	0
11	SEM24	Hit not found		0	9	1	5	5	0
12	SEM27	mediator complex subunit SOH1	3.0E-54	0.9	14	3	7	8	0
13	SEM29	Probable mediator complex subunit SOH1	0.42	0.92	12	3	6	8	1
14	SEM36	Hit not found		0.69	5	1	4	3	0
15	SEM37	unknown protein	1.0E-30	0.88	7	2	4	5	1
16	SEM38	WD-repeat containing protein	0	0.9	4	1	4	1	1
17	SEM41	Hit not found		0.85	13	8	4	5	1
18	SEM42	hypothetical protein	2.0E-18	0.88	6	1	3	2	0
19	SEM46	pi starvation-induced protein	588	0.87	7	1	3	1	0
20	SEM53	hypothetical protein OsJ_026388	2.0E-18	0.77	10	1	7	6	0
21	SEM55	zinc transporter	1.0E-49	0.73	6	2	1	1	0
22	SEM57	Hit not found		0.91	11	2	8	6	0
23	SEM58	sugar-starvation induced protein	0.001	0.84	9	5	5	5	0
24	SEM60	unfertilized embryo sac 16	1.0E-53	0.86	8	2	4	3	2
25	SEM61	zinc-finger protein	1.0E-11	0.72	12	6	2	7	1
26	SEM63	Hit not found		0.12	9	3	1	1	1
27	SEM68	Hit not found		0.7	5	6	2	4	1
28	SEM72	30S ribosomal protein S17, chloroplast precursor	0.081	0.31	4	2	1	3	1
29	SEM76	cysteine proteinase	3.0E-53	0.92	15	4	9	0	0
30	SEM78	ubiquitin-protein ligase-like	4.0E-15	0	13	2	5	3	3
31	SEM79	Nitrilase-associated protein	1.0E-11	0.91	7	3	2	1	1
32	SEM80	heavy meromyosin-like	8.0E-19	0.92	5	1	1	1	0
33	SEM82	cbs domain-containing	6.0E-74	0.82	4	3	1	3	0
34	SEM83	Dof zinc finger protein MNB1A	5.0E-12	0.88	17	8	4	7	2

Supplemental Table 1. Continue.

35	SEM84	Hit not found		0.91	10	2	2	4	0
36	SEM85	pectin-glucuronyltransferase	1.0E-37	0.91	4	1	2	0	1
37	SEM86	chitin-inducible gibberellin-responsive protein	1.0E-81	0.64	8	4	4	4	0
38	SEM90	hypothetical protein	9.0E-18	0.71	8	2	4	3	0
39	SEM92	bZIP transcription factor	7.0E-24	0.58	8	2	1	0	0
40	SEM94	Hit not found		0.82	6	0	3	2	1
41	SEM97	ankyrin-like protein	3.0E-17	0.83	17	0	2	0	1
42	SEM98	Hit not found		0.79	8	2	3	5	1
43	SEM99	Hit not found		0.9	11	1	3	2	0
44	SEM105	exocyst subunit EXO70 family protein	6.0E-46	0.92	10	2	2	3	0
45	SEM106	hypothetical protein	2.0E-10	0.91	5	2	2	3	0
46	SEM108	VP15	4.0E-30	0.66	7	3	5	2	1
47	SEM110	Cortical cell delineating protein precursor	3.0E-24	0.9	14	5	3	2	0
48	SEM112	S-adenosylmethionine decarboxylase	7.0E-31	0.92	8	3	3	6	0
49	SEM113	hypothetical protein	8.0E-39	0.92	11	6	3	5	2
50	SEM117	nuclease I	1.0E-71	0.76	6	2	1	2	0
51	SEM119	hypothetical protein OsI_018669	1.0E-35	0.41	7	3	1	3	1
52	SEM120	protein	1.0E-29	0.84	5	1	0	0	0
53	SEM124	bet v i allergen-like	4.0E-70	0.23	7	2	3	0	1
54	SEM125	hypothetical protein	1.0E-17	0.59	7	3	4	1	0
55	SEM127	Hit not found		0.92	7	2	1	3	0
56	SEM131	lipid transfer protein precursor	1.0E-28	0.85	4	4	4	3	0
57	SEM132	RRM-containing protein	1.0E-16	0.93	11	3	4	5	2
58	SEM136	Hit not found		0.93	6	2	2	3	1
59	SEM140	Hit not found		0.89	5	0	1	1	1
60	SEM141	GRC14_ORYSJPutative glutaredoxin-C14 precursor	1.0E-43	0.81	21	4	7	2	0
61	SEM143	Hit not found		0.74	3	5	4	7	0
62	SEM147	hypothetical protein OsJ_003463	1.0E-18	0.78	8	7	2	0	0
63	SEM148	Hit not found		0.91	4	3	1	2	0
64	SEM154	ribosomal protein L4/L1 family protein	8.0E-133	0.85	9	3	4	5	1
65	SEM158	Hit not found		0.72	10	2	5	5	1
66	SEM159	exhydrolase II	1.0E-18	0.91	10	4	2	4	0
67	SEM161	auxin efflux carrier	1.0E-37	0.68	6	6	2	6	0
68	SEM164	Hit not found		0.86	11	4	4	5	1
69	SEM166	Hit not found		0.62	12	5	3	4	0
70	SEM167	Hit not found		0.92	15	7	5	11	2

Supplemental Table 1. Continue.

71	SEM168	glucanase	1.0E-42	0.5	7	3	3	4	0
72	SEM174	hypothetical protein	3.0E-05	0.2	4	6	3	2	0
73	SEM176	(DSRNA-BINDING PROTEIN 3); double-stranded RNA binding	1.0E-55	0.88	5	2	1	1	1
74	SEM179	receptor protein kinase PERK1	1.0E-50	0.89	6	4	1	2	0
75	SEM180	hypothetical protein OsI_012598	1.0E-12	0.58	8	2	3	4	1
76	SEM184	positive transcription elongation factor/ zinc ion binding	2.0E-46	0.92	6	1	4	3	0
77	SEM189	auxin efflux carrier	3.0E-62	0.91	15	0	3	2	2
78	SEM190	Hit not found		0.74	9	4	5	3	0
79	SEM191	Hit not found		0.61	4	1	2	1	0
80	SEM195	Hit not found		0.44	3	1	2	1	0
81	SEM199	hypothetical protein OsJ_007772	1.0E-29	0.66	11	3	7	3	0
82	SEM200	cytochrome p450	5.0E-66	0.9	8	2	4	0	2
83	SEM203	wd40 repeat protein	1.0E-171	0.79	5	0	5	0	0
84	SEM206	lysine decarboxylase-like protein	3.0E-86	0.86	12	3	2	5	1
85	SEM207	disease resistance protein	3.0E-104	0.72	5	0	4	0	1
86	SEM208	phytoene synthase	1.0E-76	0.83	9	1	7	7	1
87	SEM211	hypothetical protein OsI_020099	7.0E-38	0.5	5	2	4	1	0
88	SEM213	Hit not found		0.68	7	1	3	4	0
89	SEM214	Hit not found		0.87	4	2	2	1	0
90	SEM215	Hit not found		0.5	5	2	0	2	0
91	SEM217	hypothetical protein	3.0E-16	0.91	6	1	4	0	1
92	SEM219	zinc finger	1.0E-32	0.37	14	3	3	4	2
93	SEM220	Chitin-inducible gibberellin-responsive protein 1	2.0E-56	0.82	6	5	4	2	0
94	SEM221	calcium-responsive transcription coactivator	1.0E-54	0.92	17	4	1	5	1
95	SEM223	DNA-dependent RNA polymerase II	8.0E-24	0.81	14	8	9	4	0
96	SEM227	GT-2 factor	0.001	0.83	7	1	2	2	1
97	SEM229	diphosphate-fructose-6-phosphate 1-phosphotransferase	0	0.65	5	3	3	4	1
98	SEM231	lateral root primordia	3.0E-40	0.86	6	5	5	2	0
99	SEM232	ubiquitin C-terminal hydrolase	1.0E-101	0.9	9	0	3	2	1
100	SEM233	gene X-like protein	4.0E-49	0.77	6	0	6	1	0
101	SEM234	family transcription factor containing protein	6.0E-11	0.79	5	0	4	0	0
102	SEM235	Auxin-responsive GH3-like protein 1	5.0E-34	0.93	11	1	4	2	1
103	SEM236	S-receptor kinase homolog precursor-like	4.0E-04	0.87	5	2	3	2	0
104	SEM237	hypothetical protein	0.011	0.12	3	3	2	5	3
105	SEM238	protein kinase domain containing	5.0E-15	0.79	13	11	4	5	0
106	SEM239	CAD1 (constitutively activated cell death 1);oxidoreductase	3.0E-21	0.83	6	1	3	4	1

Supplemental Table 1. Continue.

107	SEM246	hypothetical protein	5.5	0.64	8	3	4	2	0
108	SEM250	flavonol 4'-sulfotransferase	5.0E-13	0.62	4	4	0	2	0
109	SEM254	uv-damaged dna-binding	3.0E-92	0.88	9	6	5	7	1
110	SEM255	hypothetical protein OsI_008838	1.0E-16	0.64	9	5	4	0	0
111	SEM257	Hit not found		0.89	4	0	4	4	0
112	SEM258	TPA_exp: GRP21	7.7	0.92	10	4	2	1	0
113	SEM261	hypothetical protein	1.0E-08	0.61	7	1	5	3	0
114	SEM263	hypothetical protein OsJ_008782	1.0E-12	0.93	12	2	7	7	0
115	SEM265	Hit not found		0.85	6	0	2	0	0
116	SEM271	silverleaf whitefly-induced protein 1	9.0E-77	0.56	5	1	2	2	2
117	SEM273	cdc2 protein kinases-like	1.0E-126	0.68	5	2	2	1	0
118	SEM275	iron deficiency protein Ids3	3.0E-12	0.93	7	1	3	0	3
119	SEM276	indeterminate spikelet 1	3.0E-96	0.76	6	1	4	4	1
120	SEM279	phospholipid transfer protein	1.0E-19	0.81	10	3	2	0	0
121	SEM280	peroxidase atp8a	6.0E-147	0.89	9	2	1	3	1
122	SEM282	retrotransposon protein, putative, Ty1-copia subclass	2.0E-75	0.83	9	1	3	3	1
123	SEM285	retrotransposon protein, putative, Ty1-copia subclass	1.0E-75	0.91	5	1	1	0	0
124	SEM287	branched-chain amino acid aminotransferase	9.0E-38	0.74	11	6	1	5	0
125	SEM288	ring-h2 finger proteinexpressed	1.0E-140	0.79	12	2	5	4	0
126	SEM290	Proteasome subunit alpha type 4-1	5.0E-80	0.89	13	3	1	1	1
127	SEM294	hypothetical protein OsJ_014087	1.0E-12	0.61	5	3	2	1	0
128	SEM295	Hit not found		0.88	7	1	3	0	0
129	SEM297	g-patch domain containing	3.0E-104	0.87	11	0	6	3	2
130	SEM298	Hit not found		0.83	8	1	2	3	0
131	SEM302	Hit not found		0.72	8	4	3	3	0
132	SEM303	hypothetical protein	8.0E-23	0.9	5	2	1	1	0
133	SEM306	ccr4-not transcription complex subunit 7	2.0E-14	0	6	3	2	1	1
134	SEM307	Hit not found		0.7	7	0	4	2	1
135	SEM308	Hit not found		0.93	5	2	5	7	2
136	SEM310	Hit not found		0.89	10	6	3	4	0
137	SEM313	bicoid-interacting 3	6.0E-52	0.9	6	3	3	0	0
138	SEM314	zinc transporter	9.0E-56	0.83	8	1	5	0	0
139	SEM315	membrane protein-like	1.0E-41	0.91	8	5	0	2	2
140	SEM319	Hit not found		0.67	5	1	2	1	1
141	SEM320	Hit not found		0.5	6	2	3	2	1
142	SEM321	6b-interacting protein 1	3.0E-40	0.5	10	1	2	2	0

Supplemental Table 1. Continue.

143	SEM327	peroxisomal Ca-dependent solute carrier	1.0E-84	0.93	9	3	5	1	0
144	SEM328	Transcriptional corepressor LEUNIG	3.0E-28	0.93	7	0	4	3	0
145	SEM329	beta-1,3-glucanase precursor	3.0E-36	0.91	5	1	1	2	0
146	SEM332	rna recognition motif-containing	2.0E-89	0.72	7	3	3	3	1
147	SEM336	ubiquitin-conjugating enzyme -like	8.0E-45	0.81	5	2	2	3	0
148	SEM337	Hit not found		0.66	9	4	7	4	0
149	SEM338	nodulin21 family	5.0E-73	0.7	20	1	13	12	2
150	SEM339	Hypoxia induced protein conserved region containing protein,	6.0E-18	0.91	15	5	0	4	0
151	SEM341	Glucan 1,3-beta-glucosidase precursor	3.0E-46	0.75	15	4	3	3	1
152	SEM344	enhancer of rudimentary	6.0E-53	0.5	2	1	5	6	1
153	SEM349	symbiosis-related protein-like protein	6.0E-62	0.93	6	2	3	0	1
154	SEM350	ring-h2 zinc finger protein	5.0E-51	0.87	6	0	5	0	0
155	SEM351	Hit not found		0.92	5	2	3	0	0
156	SEM353	reverse transcriptase family member	1.0E-120	0.71	4	1	2	3	1
157	SEM355	Hypoxia induced protein conserved region containing protein	8.0E-23	0.64	6	4	1	2	0
158	SEM358	Hit not found		0.92	8	2	5	3	0
159	SEM361	Hit not found		0.93	11	4	3	6	0
160	SEM366	Hit not found		0.9	6	4	1	0	0
161	SEM367	Hit not found		0.88	6	4	0	4	0
162	SEM368	protein tyrosine phosphatase	2.0E-84	0.77	10	2	3	2	0
163	SEM369	ethylene-responsive transcriptional coactivator	2.0E-71	0.93	7	0	2	1	4
164	SEM371	phosphatidate cytidylyltransferase family	5.0E-53	0.64	8	4	4	2	0
165	SEM372	hypothetical protein	3.0E-14	0.8	13	6	7	2	0
166	SEM373	hypothetical protein OsI_028313	1.0E-50	0.88	4	1	3	1	0
167	SEM374	mitochondrial carrier protein	8.0E-50	0.65	7	3	3	1	0
168	SEM375	ETCHED1 protein	4.0E-20	0.9	11	4	6	8	2
169	SEM377	wound-responsive protein-related	3.0E-11	0.75	5	2	3	3	1
170	SEM379	hypothetical protein	8.0E-26	0.93	8	1	7	1	0
171	SEM384	Hit not found		0.86	4	0	3	2	1
172	SEM390	cytochrome p450	1.0E-57	0.92	13	7	5	3	0
173	SEM391	Hit not found		0.92	6	2	2	0	1
174	SEM392	hypothetical protein OsJ_000721	1.0E-59	0.76	3	0	1	0	0
175	SEM398	Hit not found		0.91	4	2	1	3	0
176	SEM400	Hit not found		0.93	8	0	3	3	0
177	SEM401	Alcohol dehydrogenase 2	3.0E-100	0.41	10	10	8	2	0
178	SEM403	hypothetical protein	3.0E-13	0.69	6	1	2	0	2

Supplemental Table 1. Continue.

179	SEM407	diphosphate-fructose-6-phosphate 1- phosphotransferase	4.0E-69	0.83	4	1	5	1	0
180	SEM408	growth-regulating factor 6	5.0E-72	0	6	1	3	3	2
181	SEM412	Hit not found		0.93	14	4	8	4	2
182	SEM415	Rop family GTPase ROP5	8.0E-65	0.9	4	1	2	1	0
183	SEM417	transcription factor iib	7.0E-114	0.93	28	1	3	0	0
184	SEM418	Hit not found		0.8	9	1	2	0	0
185	SEM419	serine threonine protein kinase	6.0E-16	0.91	11	2	3	1	1
186	SEM421	hypothetical protein OsI_010647	1.0E-16	0.91	6	3	3	0	0
187	SEM422	cytochrome p450	8.0E-32	0.89	8	6	4	4	0
188	SEM425	Fructose-1,6-bisphosphatase, chloroplast precursor (FBPase)	2.0E-26	0.58	10	3	4	2	0
189	SEM426	Sugar transporter family protein	1.0E-14	0.8	13	2	3	2	2
190	SEM427	dTDP-glucose 4-6-dehydratase	1.0E-50	0.91	7	6	4	5	0
191	SEM428	fructose-bisphosphate aldolase	3.0E-07	0.92	11	0	3	1	1
192	SEM430	sucrose synthase	7.0E-40	0.12	7	2	0	0	2
193	SEM432	Sugar transporter family protein	3.0E-70	0.27	11	1	4	1	0
194	SEM433	diphosphate-fructose-6-phosphate 1-phosphotransferase	3.0E-42	0.85	5	2	3	2	0
195	SEM434	disease resistance protein (TIR-NBS-LRR class)	8.0E-07	0.76	9	2	3	2	0
196	SEM435	stress-induced protein sti1	1.0E-31	0.89	9	1	2	2	0
197	SEM436	disease resistance response protein-related/dirigent protein-related	6.0E-45	0.7	8	1	0	1	0
198	SEM437	NBS-LRR disease resistance protein homologue	2.0E-16	0.83	14	3	2	1	0
199	SEM439	Pathogenesis-related protein PR-1	2.0E-19	0.2	5	2	0	2	0
200	SEM440	disease resistance-responsive family protein	2.0E-24	0.46	12	0	2	2	0
201	SEM442	disease resistance protein	4.0E-49	0.85	10	3	3	1	0
202	SEM443	dehydration responsive element binding protein	9.0E-13	0.31	12	2	1	1	0
203	SEM444	Dehydration-responsive element-binding protein 2D (DREB2D protein)	9.0E-15	0.82	13	1	5	3	0
204	SEM446	cold shock protein-1	1.0E-29	0.92	11	1	1	1	1
205	SEM447	low temperature and salt responsive protein-like	4.0E-12	0.37	3	0	1	1	0
206	SEM449	Heat shock protein 81-1	5.0E-99	0.89	17	2	2	2	0
207	SEM450	light-dependent short hypocotyl 1	2.0E-46	0.36	4	2	1	1	0
208	SEM453	cold induced protein-like	4.0E-26	0.93	5	2	3	1	0
209	SEM454	water-stress protein-like protein	4.0E-16	0.84	2	1	1	1	0
210	SEM456	Saccharum hybrid cultivar soluble acid invertase (ShinvA) mRNA	*Saccharum*	0.41	5	2	0	0	1
211	SEST3	disease resistance protein I2	2.0E-14	0.54	5	2	5	2	0
212	SEST4	early-responsive to dehydration protein	6.0E-59	0.65	19	1	1	1	0

Note: Markers highlighted with blue have been reported from other studies.

Supplemental Table 2. Details of the new 191 EST–SSR markers developed from the SUCEST database.

Marker	Cluster	Motif	Forward primer	Reverse primer	Time (°C)	Expected size (bp)	PIC	Allele number	Simplex		Double
									Phil6 607	S6	simplex
SEM1	SCEPRZ1008B02.g	(agc)10	ACCATCGCAATCGATGTTTA	AACTGGATGGCGTACAATCA	56	215	0.6	9	5	1	1
SEM2	SCACAD1035F09.g	(cg)7(ct)14	ACACCGAGCTGTCCCAAT	GCATCTGATGAGCCTGTGAA	56	164	0.9	7	2	1	1
SEM7	SCAGLB2047G03.g	(tttc)15	GGAAGGTATGGGTGCTATGC	ACAGGGCAATAACAGGGGTA	58	167	0.6	8	4	1	0
SEM11	SCAGRT2041A11.g	(tg)16	AGGGCTTGGAAGAAAGGAAT	TGGCAAGCAACAGCTAAAAC	58	190	0.9	11	4	2	1
SEM13	SCBFAM2022D09.g	(gaaa)11	GCGAGAGAAGCTAGGAAGCA	CGCAGATCCTCTTGAACCTC	62	161	0.7	7	2	2	0
SEM14	SCBGAM1091A05.g	(ta)10(cag)5	GAGCAACGAGCTGAAAAGTG	TGTCCTGACCTAGGATGTGC	59	245	0.8	10	2	4	1
SEM15	SCBGLR1119D12.g	(ta)26(at)5	TGTCCACAATTTTGGCTGAT	GTTGCTTGCCTGATCATTGT	56	200	0.5	5	0	2	0
SEM16	SCCCAD1003H03.g	(aag)15	CCTTCCTTGGCCTCTTCTCT	TGCTGGTCGCAGTACTTGAT	58	248	0.9	3	0	0	0
SEM19	SCCCCL4002A08.g	(gca)5(caa)13	CAGCCCATTAACCAAGCAAT	GAAGCAGCTGTTGCTCACTG	58	184	0.9	13	5	2	1
SEM20	SCCCCL4004D08.g	(ataga)12	CCGGCTGTGAAAATTAGGTT	TCGAATTGGTCAAGACTCTCC	58	219	0	6	2	2	0
SEM24	SCCCCL4015G07.g	(ct)10(ac)5	CAATTCGTGGCTTGTGTTTG	AGCAGAATCGGCAAGGTAAA	58	197	0	9	4	1	0
SEM36	SCCCLR1078G05.g	(ta)12	CCATGTGCAGCATTTAACAA	TGGACATGCTAATGACTACTGC	56	152	0.7	5	3	0	0
SEM37	SCCCLR1079D03.g	(ga)15	CTTCCTGCTTCGAACATTTG	ACGAGGTAGATCCCGAAGG	56	214	0.9	7	3	2	1
SEM38	SCCCLR1080G12.g	(ca)10	TGAATTGCACAAAACACCAA	GACGGTGTAAACAAGCTGTGA	55	249	0.9	4	1	3	0
SEM41	SCCCRT2001B06.g	(ag)24	CCCCCTTGACACCTCTGTAT	TAGAACGAACCAGACGACCA	59	248	0.9	13	4	8	1
SEM42	SCCCRZ2004G11.g	(aag)7(ta)15	GACTTCGGGAAGAAGGAGGT	ACCAAGCACATCCAGCAGTA	59	285	0.9	6	2	0	0
SEM53	SCEPAM1020C01.g	(ct)20(ct)20	CCGCCTTCTCCTTAGTGACA	CCACAAGCCTAATACAGCTCAA	62	192	0.8	10	6	0	0
SEM55	SCEPAM2057B09.g	(ct)15(ct)5	ACGGCATCAGATTCAGATCA	ATGGCTTTCCATCTCGTGAC	56	277	0.7	6	1	2	1
SEM57	SCEPRT2048F09.g	(cag)11	TCCAGAAGTACGTGGAGACG	ACGACAGCAGGTCGAACAT	56	222	0.9	11	4	2	0
SEM58	SCEPSB1130G10.g	(cacta)15	CCAACCAACCTCGACATTCT	CCATGTGATCTGACCTGGTG	60	209	0.8	9	4	2	1
SEM60	SCEQAM1041B03.g	(ag)12	TGCTAACACATTTCAAGAAAGAGA	GATCCAATCCGAGGAAAAGT	58	155	0.9	8	4	0	3
SEM61	SCEQAM1041H02.g	(ct)12	GTTCAGAACACGTGCAGCAT	CACGCTTGACATGAGAGGAA	59	165	0.7	12	3	5	2
SEM63	SCEQHR1078A09.g	(ag)25	GGTCGGTGCTCTGTTCTTTT	CCTGCAGCAGAGACGAGAT	59	197	0.1	9	1	0	2
SEM68	SCEQRT1032H07.g	(ta)6(ac)11(tc)6	CCCTGAGGTCTCTCTCCACT	TGCCATAGGACAAGAGTTTAACA	64	239	0.7	5	1	3	1
SEM72	SCEQSD2075F10.g	(ag)14	GAACCTTATCGGTAGCCTCCT	GAGCGCCATAGGAGAAGTG	58	158	0.3	4	1	2	1
SEM78	SCEZRT2023F09.g	(ggt)5(atag)14	GTGGTCGCAGACGAGGTC	CTCCGCATTAGCCATTTCC	58	222	0	13	3	0	3
SEM79	SCEZRT2024C04.g	(ac)13	GATGGAACAGATGCGACAAG	GTTCATCGTAACCTGCTGGA	58	207	0.9	7	1	2	1
SEM80	SCEZRZ3015G05.g	(ta)27	GCAGATGAGAGGGCAAAAGT	CGCCTGCAGATGAATCATAG	59	244	0.9	5	0	1	0
SEM82	SCEZRZ3017G04.g	(ga)12	AGTACAAGGCACAGCCAGAG	GGACATGAGGTACACCCAGA	62	197	0.8	4	1	3	0
SEM83	SCEZRZ3096G10.g	(atg)10	TCCTCCTCTTGTTGCAGTTG	GTCGTCGTCACGATCATCTC	58	184	0.9	17	1	6	2
SEM84	SCEZSB1094A08.g	(tc)10	TGTAGCAATTCCTTGCGTTG	CAACAAATACAATGCCAATCG	58	229	0.9	10	3	4	3
SEM85	SCJFAD1011F07.b	(ga)10(agag)5	CACCTAGTGAAAGGGGCAAA	CCTGAAGCCTTGGTAGCATC	59	242	0.9	4	2	0	0
SEM86	SCJFLR1035E04.g	(ga)10	CGAGAACTAGCATAGCACAAGA	AACAACTGGTGCAAGTCCAT	58	172	0.6	8	1	2	1

Supplemental Table 2. Continue.

SEM90	SCJFRZ2010C01.g	(ag)17	CGTGGGCTAATACAGCTCAA	CCCACTGGCACAATCTTCTT	60	190	0.7	8	0	3	0
SEM92	SCJFSB1010B12.g	(tc)10	TGTGCCGTTGCCTAATAACA	GGCAAGCTTCCTCAGTTTCTT	56	209	0.6	8	0	2	2
SEM94	SCJFST1014E07.g	(tc)10	TTATCTCTTCCCTGCGGTCT	CTGGCTGGGGAGCAACTT	56	196	0.8	6	2	0	1
SEM98	SCJLRT1019D02.g	(tg)15	GCCAGCAGAATGCTTAACAA	TCTTACTCTGTCCCCCAACC	58	233	0.8	8	3	2	1
SEM99	SCJLRT3076A02.g	(ta)12	TTACCTCCGGCAACGTTAAA	TGCAGGCATATGGTAGTCCA	56	175	0.9	11	5	0	0
SEM105	SCMCAM1100G01.g	(ctttt)18	AGGGGCCTCAAGTTGTTCTT	CGGTCTCATGGTCACCTTTT	58	220	0.9	10	1	5	2
SEM106	SCMCCL6027C07.g	(tc)24	AGGTTGCTGATGGTCCTCAC	CAAGAAGGGAAGCAGGGACT	62	197	0.9	5	1	2	0
SEM108	SCMCFL5008F03.g	(ag)10	CGACTTGTGTGGAGGTGAAA	TGGATCAATGTGAACAAAATCG	58	164	0.7	7	2	2	0
SEM112	SCMCST1057D03.g	(tc)14	CCTTCTGCAGACGAGTTGAA	ACCTGACCAGCAAATCAACA	58	159	0.9	8	3	3	0
SEM113	SCPIRT3024F01.g	(ga)17(gga)5	TTCCGGTTTACCCTGCATAG	TTCCTCAGGGCCCTTTTATT	58	232	0.9	11	2	3	2
SEM117	SCQGLV1018G08.g	(at)41	GCGTGGCACTGACTACAAGT	CAATGTTCTGTGGTCTGCACT	56	246	0.8	6	3	1	1
SEM119	SCQSHR1020F04.g	(ct)16	CCGCGTGCTCTCTCTCTCT	ATTGCCATCACCTCATGCTT	56	206	0.4	7	1	2	1
SEM124	SCRFLB1055F01.g	(tcc)11	CCTTGATGTGCTTGACGAG	CCAACGAGCAAAAGTAAACG	56	270	0.2	7	3	0	2
SEM125	SCRFRT3058D07.g	(ct)16	TATTCTCTCCGGATCCCCTA	ATTCAAAGCGCAACACAGTC	58	160	0.6	7	2	3	0
SEM131	SCRUFL4024B04.g	(ac)18(cgc)6	AAAGGAAAGCAAACCCAAGG	GCAGTCGTTGTCGTAGCAGA	58	245	0.9	4	0	1	2
SEM132	SCRURT2010E12.g	(ct)18	CCCTCCACCTCTTTGCTC	TAGAAAGACCTGCCCTCCTG	58	202	0.9	11	3	2	1
SEM136	SCSBAM1085C09.g	(cgg)12	CTCTGACCCGAGCAAAGG	GCCAATAAACAGCAGGGGTA	56	238	0.9	6	1	2	1
SEM140	SCSFAD1114H02.g	(at)23	GTGTTTTGGAGACCGTGTCA	CGATTGTTGCGCTGTACATC	58	127	0.9	5	1	0	1
SEM141	SCSFFL4083B01.g	(ac)16(ag)13(gt)5	ATCATCCACAGCTAGCAGCA	GGTTTTGCCTTGGTTTTTGA	56	174	0.8	21	6	6	2
SEM143	SCSFHR1045G08.g	(ct)9(ct)12	CCCCTCTCCCTCAGTCTTCT	CCATGCTGTCAGGATCCAC	60	279	0.7	3	1	1	1
SEM147	SCSGHR1070F11.g	(tcc)10	GCCTCTTCCTCCTCCACTC	GACGACCGTCCTTGTTGAG	58	250	0.8	8	1	2	0
SEM148	SCSGLR1025D03.g	(tca)12	GCTACCGGATGGATAAAAGC	CTGACCGAAATGATCAAGGA	56	245	0.9	4	1	3	0
SEM154	SCUTCL6035D02.g	(ct)14	ACCGAGGTAGGAGGGAGTGT	GCTCGCCATGAATAGAAAGG	58	219	0.9	9	4	1	1
SEM158	SCVPCL6044A06.g	(ag)12	GGATGGTTAAAGCGGAAACA	GGAAACAGTGTACGCCCAGT	56	151	0.7	10	4	1	1
SEM159	SCVPHR1089A09.g	(ag)13	CTGGTGGAATAACTCGCTGA	CTCAAGGCAAGAACGAATCC	60	279	0.9	10	4	3	1
SEM161	SCVPHR1094C01.g	(agg)10	GAACTGCTCACTGGCTCCTC	GTAGAAGTCCGTCGCCGTAA	58	186	0.7	6	1	1	1
SEM164	SCVPLR2019H04.g	(ag)18	GAGGTTATGGGGAAACCAGA	GATTGCAGCCGTAAACTTGA	56	211	0.9	11	1	4	1
SEM166	SCVPRT2080G09.g	(gt)22	AGCGCATCTTGCTTATTTGA	ATGCATGATCATCGAGGAAG	56	171	0.6	12	4	1	1
SEM167	SCVPRT2083D03.g	(gt)30	CGGATCTTGGCTCCTTCTCT	AGCCTTGATTGGCAATGGTA	58	223	0.9	15	4	7	2
SEM168	SCVPRZ2037E02.g	(ct)11	AACGTCAGCCGCTACAACTT	CTTCCCTTTTGCGAAGAAAA	56	143	0.5	7	2	2	0
SEM174	SCQGFL3059G12.g	(ta)11	CTCACCGCAGCTCTTTTTCT	CACAAGCTATGCGGTCAAAA	58	249	0.2	4	0	1	0
SEM176	SCRFFL1029H06.g	(cctc)5(ct)10	CGCCATAACCATAACCACAG	CCTCCCTCCGCTACTTCCTA	58	188	0.9	5	1	2	0
SEM179	SCSBFL1101G01.g	(aaag)10	TATTCCACCGGGAACAAGAA	GGGATTGTAGCGACGAGTTG	58	210	0.9	6	0	2	0
SEM180	SCSBFL1104E01.b	(ct)12(tgc)6	TTCCACATCAAGCAAGCAAG	ATGACATCAGGAGGGAGACC	58	206	0.6	8	2	2	1
SEM184	SCVPFL1073A11.g	(ga)10(ggc)7	ACCAACGCGACGAGAGAG	GCCTGAACTGGTCGTAGGTC	58	206	0.9	6	3	0	0
SEM189	SCJFAD1013E12.g	(agg)6(ggaa)8	GAACTGCTCACTGGCTCCTC	GTAGAAGTCCGTCGCCGTAA	59	209	0.9	15	4	3	2
SEM190	SCRLAD1098A04.g	(ct)9	CTAGCACGGCAATACAGGC	AGATCTGTTGGGTGCTCGTC	59	189	0.7	9	2	3	0

Supplemental Table 2. Continue.

SEM191	SCRLAD1138A05.g	(gcc)8(cca)5	CCAGTCGCGATTCTTCCAC	AAGGGACGGGGAGAAAAATA	55	176	0.6	4	2	0	1
SEM195	SCSGAD1008F08.g	(ggc)9(gag)5	CTTCCCGTCGCTCTTACCT	CTCCTCCTCCTCCTCTCCAC	60	183	0.4	3	0	1	0
SEM199	SCEPAM1021B02.g	(tg)9	CTCTCGAGGAGGTGGATGAG	CTGCAAGTTTGTTGGCTGAA	56	237	0.7	11	6	3	1
SEM200	SCEPAM1050A03.g	(tc)9	CTGCAGGATCACCTGGAAC	TAAACCCACGCTGACAGACA	56	238	0.9	8	2	1	4
SEM203	SCEQAM1036D03.g	(cgc)6(cgc)8	GCGGCCTCATACGTGTAGAT	TCTCTTCCCCTCACCAGAAA	56	237	0.8	5	3	0	0
SEM206	SCVPAM1056A04.g	(tctt)9	CATGGTAGCTCCGCTTCTTC	GCGAGAAGCTAGGAAGCACA	59	197	0.9	12	2	3	1
SEM207	SCACAM2043G03.g	(gct)8	GGCACACCTCGAGAGACC	ACTCCTCCTCCTCGCTTAGG	60	151	0.7	5	3	0	1
SEM211	SCCCAM2001E04.g	(cgc)8	CGGTCGTCTCTTCCTCCTC	CTACTACCACCCGGACCAGA	59	212	0.5	5	1	1	1
SEM213	SCCCAM2C08B11.g	(tc)5(tc)5(ct)5(ct)6	CTCTCCGACTCGTCTTCCAC	GCGGACTGCAAAAGAGAGAT	56	241	0.7	7	3	1	0
SEM214	SCEQAM2037C11.g	(cag)8(cg)5	ATCGGCTCCAGTCAGAGAGA	CCTGGTGAAGGCTCATGATT	59	323	0.9	4	1	2	1
SEM215	SCEZAM2033H10.g	(tc)8	GCCGAAGAGGAATCTACGAG	GTTTGTCTTCCTCCCTGTGC	56	193	0.5	5	0	3	0
SEM217	SCEZAM2096F07.g	(ga)9	CACGGGGAGACGAGAGAC	CCAACAACAACCAGAATATCG	54	174	0.9	6	3	0	1
SEM219	SCMCAM2084A04.g	(cag)6(cag)5(gca)5(cag)5	AAGTACGGAGCGCAGTGTAG	ACCGCCTTGTACTCCAAATC	56	228	0.4	14	3	1	2
SEM220	SCMCAM2084F10.b	(at)8	AAGCTCCTTGCCTGCTACTC	CAAAGGGCATCCTTTCTGAT	55	218	0.8	6	1	2	1
SEM221	SCQGAM2028B01.g	(ccg)8	GCCTCTCTCTGCTCAGCCTA	CTCCTCATCTCTCGCCAAA	56	170	0.9	17	5	1	2
SEM223	SCSGAM2076E10.g	(tc)8	CACAGCACTTGCCAAGCTAA	AGTTTCACAAAGGGCGACTG	56	216	0.8	14	7	4	0
SEM227	SCCCCL3005D03.b	(ct)8	GCTACAGTGCCTCTCCCTCT	CTAGAAGCAGAAGTGGAGTGCT	59	287	0.8	7	2	1	0
SEM231	SCCCCL4013B10.g	(cgt)5(gcg)9(cg)5	CCGTTCTACACCTCCAACAT	GACCGTGACCATCTGCTG	57	426	0.9	6	1	2	0
SEM232	SCCCCL4014F09.g	(ga)8	CAACTCCAGCTCCAGTCTCC	CTTTTCGCGAAGTGAACACA	58	311	0.9	9	4	1	0
SEM233	SCCCCL4015B01.g	(tgt)8	TTGCTTGGGACAAAAGGCTA	ATCTTGCAAAGGAAGGAGCA	55	336	0.8	6	3	0	0
SEM234	SCACCL6009D08.g	(acc)8	GGACATGCTGCTCCCTACAT	AGGAGGACTGGTGGTTGAGG	60	211	0.8	5	3	0	0
SEM235	SCACCL6010C05.g	(tc)9	CATCGGCTCATCATAACGAA	AGCTACTTCAGCCCCAAGTG	55	250	0.9	11	5	0	1
SEM236	SCCCCL6003H04.g	(ct)9	CCCTTTGCTTCCCCTTTACT	GAGGCGCCTTACTGTTCTTG	56	193	0.9	5	2	1	0
SEM237	SCEPCL6023G01.g	(gca)8	AGGGAAAGAGACGAGGGAGA	CGTATCTCCGACCACTCCAC	59	167	0.1	3	1	1	0
SEM238	SCEPCL6029D06.g	(gcg)8	CTCTCCCCCAACTCTCTCTG	TCCGACGTCAACGTCTCAG	59	176	0.8	13	3	7	2
SEM239	SCRLCL6030D09.g	(ct)9	CGAGAAACCGTGTCCCCTA	CCCTCTCCCTCTTCCTCCT	59	155	0.8	6	1	1	0
SEM246	SCJLFL4097F08.b	(cg)5(gaaa)8	AATCGATCTTAGGGCCGGTA	ACGCCGACGAGTGAGGAC	58	276	0.6	8	2	3	1
SEM254	SCBGFL5080G03.g	(cgg)8	ACCTTACAGAGCCCACTGCT	TCGCGATAATGAGATTGAGC	59	152	0.9	9	4	4	1
SEM255	SCCCFL5062D10.g	(ag)8	CGGCGTCCACTGAAAGAG	CAGCCTCGAGTTGGGATG	56	178	0.6	9	1	0	0
SEM257	SCEZFL5084A01.g	(tg)8	TGCTGGAGACGGAGTAGCTT	ATCAGGCAAGCACACAATCA	57	159	0.9	4	1	1	1
SEM258	SCEZFL5091D04.g	(cac)9	GGAAGAGGAGGCTTCGAGAT	CTGGATAATCACGCCCAAAT	55	343	0.9	10	1	4	1
SEM261	SCAGFL8042E05.g	(cgg)8	CCATCCATCCTCTCATCTCC	AAGAGTGCTTGAGCGGATCT	56	187	0.6	7	4	1	1
SEM263	SCRLFL8053B05.g	(gga)5(gca)8	AGCCTCTGACGCTAAGATCC	CACACGCTGCAGATGTTGTT	56	208	0.9	12	7	0	0
SEM265	SCAGHR1018C11.g	(cca)8	ACACTAGCTAGCCAGCCACA	GAAGCGAGGCTATGGCTATG	57	163	0.9	6	0	0	0
SEM271	SCJFHR1034E09.g	(ccg)9	AGCAGATTCACTTCGCCACT	CGATGAGCTTGGAGAGGAG	55	157	0.6	5	2	0	2

Supplemental Table 2. Continue.

SEM273	SCQSHR1022B03.g	(cat)9	TTTCTTTTCGTCACACCCAAT	ACTCCCGTCACTCACCTGAC	55	180	0.7	5	1	2	0
SEM275	SCRUHR1074E09.g	(gag)9	TCTCATCGGATTCACACACA	GGGCAGCTTCGTAATGGT	55	242	0.9	7	1	0	1
SEM276	SCSFHR1043F12.g	(tg)5(tg)8(ag)5	AACCCGTTCTTCTTCCCCTA	CAGAGGGAGATTTGCCATGT	55	241	0.8	6	1	1	1
SEM282	SCQGLB1038F11.g	(gcc)9	GAACCTCGCAGTCTTCACAA	CACTACCTGCCTTTCTCTCG	56	191	0.8	9	4	1	2
SEM285	SCVPLB1020B05.g	(gcc)9	TCCTTGAACCTCGCAGTCTT	CTACCTGCCTCTCTCGTTCC	54	192	0.9	5	1	1	2
SEM288	SCACLR1057E07.g	(ga)8(gct)5(ctg)7	TCCGATCACAATCACAGACC	GCTGCAGCAGATGACAAACT	58	227	0.8	12	1	1	4
SEM294	SCCCLR1066D07.g	(cca)5(ccg)8	CCATACCCTGTACCGTACCC	GATGCTTGCATTCATCCTTG	55	184	0.6	5	2	3	0
SEM295	SCCCLR1066F12.g	(tc)8	CACCTCCCAGACTCTTCTCC	GTGACACCATGGTCCTGAAG	56	176	0.9	7	2	1	3
SEM297	SCCCLR1075D10.g	(cgc)8(cac)6	CACCAAACAGACTCGCATTT	CGGATCGAACTCTGTGACAT	54	214	0.9	11	5	1	1
SEM298	SCCCLR1076A04.g	(ca)8	ACGCGAGAGGGAGAGAGATA	GTCAGCAGCACGAACAGC	59	163	0.8	8	3	2	0
SEM302	SCEPLR1008H10.g	(ggc)5(cgg)8	GCGGTTTCTTGTTTTCCTTC	ACCACGACCTCGATCTCAAC	56	282	0.7	8	3	4	1
SEM303	SCEPLR1030D11.g	(agg)8	CGAAAACCCTCAAACCCTAA	CTCCTCTAGCTTCCGCTTGT	55	193	0.9	5	0	3	0
SEM306	SCJFLR1013A08.g	(cag)8(cag)9	ACCACCAATACCACCACCAC	TCGACGTTGGACTTGAGAAG	56	247	0	6	1	2	0
SEM307	SCJFLR1074A10.g	(ac)5(ca)9(ag)7	CAAACTTTTGCCCGATAGGT	CGGAGCATACCAAGTGAAGA	54	244	0.7	7	4	0	1
SEM308	SCJLLR1101F02.g	(gt)8	TCTCGACTCCCCTAATCACC	CGGACAGAAAGATCGCAGTA	55	242	0.9	5	1	1	2
SEM310	SCQGLR1019C10.g	(ga)8	AAGAAACCAACCCTCAAAGC	GTAGGGTAGCGCTGGGTAAT	55	230	0.9	10	2	4	1
SEM313	SCSGLR1084A02.g	(gcg)9	GAGGGAACACATCCCTTCTC	GCCGTAGATGAAGACCTCCT	56	211	0.9	6	3	3	0
SEM314	SCVPLR1049G12.g	(ct)8(ct)5(cgc)7	GAATATAACCGCCACCTTGC	TGGCTTTCCATCTCGTGACT	58	250	0.8	8	2	1	1
SEM315	SCACLR2007A01.g	(ca)5(ag)9	GAGGTCCTGGGAGAGACAGA	GTCTGGCCCGTAAGCTGT	60	152	0.9	8	2	2	2
SEM319	SCAGLR2026C05.g	(cgc)8	ATCGTCATCGCAAAATGC	CAACCGGAGGCACTGAGTA	56	150	0.7	5	2	1	0
SEM320	SCCCLR2002F05.g	(ag)9(ag)7	GAGGCAGCTCGACGACAC	GTCAGCTCCGCTCCTGCT	62	111	0.5	6	2	2	1
SEM321	SCQGLR2032D06.g	(ga)8(gc)5	GTCCGTCTCCACTCGAAAAC	GCGGTTGAGGTCGAGGTAG	59	222	0.5	10	4	1	1
SEM327	SCCCRT1001G10.g	(cct)8	CTCCCCTCTCGCTCATCA	AGGTTGACGATGGTGGTGAC	59	196	0.9	9	1	3	1
SEM328	SCCCRT1003H03.g	(ct)16	TCTTGCCTGTTCGTCTTCCT	ATTCCGATTCCGATTCCAAC	55	239	0.9	7	4	0	0
SEM329	SCEQRT1025C10.g	(cgg)8	CACCCAGCTCAAGTACAGCA	GCCTGTAAAAGCCTCCTGTG	59	222	0.9	5	1	1	0
SEM332	SCJFRT1009B09.g	(ggc)8	CCGCAAGGAAGAACACCTT	GCAGTGGAAGTCGACGTAGG	56	232	0.7	7	1	2	0
SEM336	SCJLRT1006C08.g	(at)5(aag)8	GCCAGGGTTCTTCAAGTGAT	TTCGTCATAGCCATCGTCAT	55	155	0.8	5	2	1	0
SEM337	SCJLRT1013F12.g	(ga)8	AGCAATGGTACGCACAAGAG	TTGCTAGTCGTCGTTCTTGG	55	202	0.7	9	1	1	1
SEM338	SCJLRT1018G02.g	(ga)8	GATCGGATCGAGAGGAGTTTT	ATACGACGAGGACGAAGTGG	56	216	0.7	20	11	1	1
SEM339	SCJLRT1019C06.g	(ag)8	AAGCGAGCGTACACCAAATC	ACGGCTCAGATGGTTGAGAG	58	163	0.9	15	2	2	0
SEM341	SCAGRT2041D09.g	(cgg)8	GTGGTTTGAGTACGCTCGTG	AGAGGGATGGCAGTATCCAG	56	249	0.8	15	5	3	1
SEM344	SCEPRT2047A05.g	(ct)15	CGTGCGCTCTCTCTCTCTCT	ATTTTGAGATGGCTGCATCA	57	171	0.5	2	0	1	1
SEM349	SCEQRT2099E08.g	(gca)8	CGAGAGGCCTTCTCTCTCTG	CGCTGACGTAGTCCTGGTAG	59	180	0.9	6	3	2	0
SEM350	SCJFRT2057F04.g	(gca)9	CCAATGGAGACGACACTCCT	GCGGACGTAGATGGAGAAGA	59	228	0.9	6	5	0	0
SEM351	SCMCRT2085E08.g	(tg)8	CGACTGTGGGAGGAGTTTGT	TTGCAGCAGTTGCTAGCTGT	57	119	0.9	5	2	1	1
SEM358	SCAGRT3048C12.g	(gaa)9	CTGGCCTCAAGAGGAAACTG	ACCAACCTCTTGACCAGCAC	59	124	0.9	8	5	2	1

Supplemental Table 2. Continue.

SEM361	SCCCRT3001D09.g	(ct)9	GTAGCCGTGGAGCATGAAGT	CTGCTGCCATTAGGAGCAAT	59	173	0.9	11	3	3	2
SEM366	SCBFRZ2045C02.g	(ca)9	CCACCTCTTCTGCCAAGAAC	CATCTTAAACTCCGGTCCACA	55	167	0.9	6	0	2	0
SEM367	SCCCRZ2001C02.g	(ag)8	AGTCAGCATCCATCCAGTCC	ATTTCTCCTGCCCTCCTCTC	59	196	0.9	6	1	0	0
SEM368	SCCCRZ2004C05.g	(gcg)9	AAACCCTCGCCTCCGATT	CCCAATGGTACCAGCAGAGT	59	241	0.8	10	3	1	1
SEM369	SCJFRZ2015A10.g	(ga)8	CGCTTCCATATCTTCTTCTTGG	TGACTCTCCGGTCCCTACAC	55	123	0.9	7	1	0	0
SEM371	SCJFRZ2034B06.g	(tg)9	GGAGAAGCATTTCAGCAACC	CCCGCTTTTCCTCTTTCTTT	54	238	0.6	8	1	2	1
SEM372	SCVPRZ2036E01.g	(at)8	GCCAAGCTAAATAGCTGCTG	ACCACCGTTTCTTTCCTGAC	57	205	0.8	13	4	4	0
SEM373	SCVPRZ2038E05.g	(ccg)8	GCGACCAAATCTGCCGTAT	CATGTAGTCGAGCGCAGAGA	58	189	0.9	4	2	1	0
SEM374	SCCCRZ3003B01.g	(ag)9	GCCTCCTCCCTCCTCTTCTA	GACTGGCTCGGAAACCCTA	59	141	0.7	7	0	2	1
SEM375	SCEPRZ3128D05.g	(ct)6(tc)9	ATGGAGGCTCGTTGTCTTTG	CCGTAATCGCCTCCACTAAA	55	174	0.9	11	4	4	0
SEM377	SCEQRZ3020E12.g	(gcg)8	GGAGAGGACGAAACCCTAGC	CGCATTGAACGCAGTTTCTA	55	233	0.8	5	3	0	1
SEM379	SCJFRZ3C03A08.b	(ctgtg)9	ACGAGGCCACCATAGAACAT	GCACAAGGTGATTGTGCTGT	56	221	0.9	8	5	1	0
SEM384	SCUTRZ3103F01.g	(cgg)9	TAGTAGCAAGCGAGGCGATA	GTCTGTTGCCTTTGATCGTG	55	228	0.9	4	2	0	0
SEM390	SCSGSB1005B12.g	(ag)9	GGGGAAGTAAGTCTCAGGTCA	GCCACCACCTCCATTATCTT	57	116	0.9	13	6	2	3
SEM391	SCUTSB1033C02.g	(ag)8	GTTCAGACTCGCGTGTTTTT	GCTGAGAACCCTTCAGCTCT	55	104	0.9	6	2	3	0
SEM392	SCUTSB1075H09.g	(ta)8	TCATGCTCACCAGCAAAGAC	TCCCGATCAGTGTGTAGACG	55	234	0.8	3	1	0	2
SEM398	SCEPSD1006D03.g	(ta)9	CGTGCAAGCTCCAATATGAT	TGCCACTGTATAGCAGCGTA	54	184	0.9	4	0	2	1
SEM400	SCEZSD1081A02.g	(ccg)8	CAGCTCATCCTCGTCAACCT	CTCCTCTGCTCCTTGTTGCT	59	225	0.9	8	5	1	0
SEM401	SCMCSD1059G09.g	(ct)9	GCTCCATTCATTTCCTCCTC	TTCGATCGATTGATGGTTGA	53	111	0.4	10	1	2	0
SEM403	SCEQSD2077B12.g	(cga)8	CCTGCATCAACCTCTCCAC	GAAGGCGAGAGAGAAGATCG	55	242	0.7	6	2	1	3
SEM408	SCJFST1048G04.g	(ga)9	CAGAGCCAGCCAGGTAAAAG	TCATCGTGTGCTGCTGGT	58	228	0	6	2	1	1
SEM412	SCJLST1022C09.g	(ct)8(ga)7	CAAGGCTGCTTCTGGTGTC	CCTCTTTGGGTTCTCTGCTC	58	246	0.9	14	3	1	1
SEM415	SCMCST1050H06.g	(tc)8	CAGCAGACGAGACGAGAGAG	AGGGTGATGAAGGGAATGAG	56	163	0.9	4	2	1	0
SEM419	SCSFST1066E06.g	(gcc)8	TGCGTGGTTGATTGAAGAAG	AGAAGCCTCTTCTGCTGCTG	60	199	0.9	11	2	3	1
SEM421	SCSGST1069F04.g	(gga)5(ctc)8	CACCCTGCTGGTCTCCTC	TCGACGTCGTGTAGTGAACC	59	170	0.9	6	1	3	0
SEM422	SCSGST1072B03.g	(ag)8	GAAGAGTGGGGACGTCTCAG	GCCAGAGGATGTGGTAGAGG	59	199	0.9	8	3	2	1
SEM425	SCSFAD1070E12.g	(gcc)5	GTGCCACCAGCAGCAAT	TCTCGTAGCTGCTCGACTTC	56	244	0.6	10	1	4	0
SEM426	SCVPAM1059C01.g	(at)5	TCGAGAGCGGTTTCATCTTT	CTTTCCTGTCAGCCAAGTGA	56	471	0.8	13	2	3	1
SEM427	SCSBFL1105H11.b	(ca)6	AAGTAGCGGAAGCATTAGTTCA	CCAAGTTCCTCCTCACCAGTA	57	277	0.9	7	1	1	1
SEM430	SCRUSB1064F09.g	(cgg)5	TCCGACTACCTCAAGTGCAAG	GACGGCATCTTCTTCTTCTCC	55	224	0.1	7	1	0	1
SEM432	SCJLST1019B07.g	(gc)6	CGCGTCCGTAGATTAGTAGCTC	AGCGAGTAGATGTTGATGACCC	56	195	0.3	11	2	2	1
SEM433	SCSBST3094H07.g	(cga)6	GACACGCCCAAAGGAAAAG	GAGATCCGGACACACATGG	54	245	0.9	5	1	1	0
SEM434	SCEZLB1007E12.g	(ta)7	TTCTTGCTTCTTTCTTTCCGTC	TCAAATCGTGCTTGCTTGAG	52	236	0.8	9	0	1	2
SEM435	SCQGLR1041A05.g	(ga)5	AGGCTGAGAGAGCAAAGAAAGA	CCTAGGATCCTTCGGGTTTC	55	164	0.9	9	2	1	3
SEM436	SCJLRT1021D04.g	(tcc)5	GGTCCCATACATAACACAAGCA	TGCATGAAGAAGCTCAGGTG	57	248	0.7	8	5	1	0
SEM437	SCQSRT2031C10.g	(tc)5	CCTGGTTCCTGCACTTGTCT	CATCACTTGCCATCTGCATT	57	217	0.8	14	7	0	1

Supplemental Table 1. Continue.

SEM439	SCACSB1117C07.g	(cgg)6	CGTCAAGCTGTAGTCCGAGAG	CTCGTCCCAGACCAGGAG	59	197	0.2	5	0	3	2
SEM440	SCACSD1018E05.g	(gac)5	AGCAACCTAATCACAGCAACAA	CCATCATCCGATCATCCTTC	56	229	0.5	12	0	2	3
SEM442	SCMCST1057C10.g	(gct)5	CATTTATTTGCCACCTAGAAGGG	AAACAGAAACCGGACAGCAC	56	195	0.9	10	2	2	1
SEM443	SCRLAD1043B06.g	(ggt)7	GGAATGGGAACAGCCACTAAC	AAGAAGGCTATCGAGGTGGG	55	323	0.3	12	1	3	3
SEM444	SCBGAD1027C03.g	(ggc)7	CACGGTTCTCCTGCTGAAAG	GACGGGGTTGTTGAAGGTG	55	313	0.8	13	1	3	2
SEM446	SCCCCL3001D10.b	(ccg)5	GAGCAGTCCCTTGCCATGT	GCCGTCGAGTACACCGTC	59	389	0.9	11	0	1	0
SEM447	SCEZFL5083C02.g	(gc)5	TGAGTTCAGTTCCTTCCCC	AGAACTCCAAGGAGCAGCAG	56	300	0.4	3	1	0	1
SEM449	SCEZLB1006B07.g	(gcc)5	TGGTGTGAGTTAGTGCCTGAGT	TAGAAGGTGTTGATGATGAGCG	55	265	0.9	17	2	2	2
SEM450	SCEZLB1007E12.g	(ta)7	TTCTTGCTTCTTTCTTTCCGTC	AGATGAACACATAGTTGCACCG	56	189	0.4	4	0	0	2
SEM453	SCBFRZ2045E11.g	(ggc)5	AGCGACATGAGCTACCGTCT	TAGTACCGCGACAGACCTTTCT	58	287	0.9	5	0	0	1
SEM454	SCSBRZ3122D09.g	(gga)6	GTAACTAGCAGCAACCCTAGCC	ATCCTCTTTTGCCTCCCCT	55	387	0.8	2	0	1	1
SEM456	AY302083	(tgc)6	TCGTCCTACAACCACGACTACA	GAGAGGCAAGCAAGGAAAGAT	56	164	0.4	5	1	1	0
SEST3	SCSFSB1097B02.g	(ta)8	CCCCGAAGATCAAGGATAGG	CGCATCTCAAATGGGAAAAT	56	413	0.5	5	0	3	1
SEST4	SCRLAD1040D08.g	(at)5	CAGGCACTGATGTCATGGAT	GAACTACACTCGCCGCTCAC	56	313	0.7	19	0	6	3