RESEARCH ARTICLE

GENIC SSRS FROM *SACCHARUM OFFICINARUM* IN WRKY AND STRESS TRANSCRIPTION FACTORS TO USE THEM AS MARKERS IN SUGARCANE

R.M. Shanthi[1]*, Ravinder Kumar[1], Padma Nimmakayala[2] and O.U.K. Reddy[2]

Abstract

SSR markers proved to be valuable in sugarcane genome analysis, enhancing the efficacy of molecular breeding programs. Genic SSR markers represent functional molecular markers as they belong to relatively conserved coding regions of the genome. A study was undertaken to mine the microsatellites in the available sugarcane unigene datasets for understanding their abundance and determining the extent of allelic diversity for WRKY and other stress factors. A high number of repeat units were observed in the open reading frames(ORFs) rather than non-coding untranslated regions (UTR's). Among trimer motifs, GCC/GGC, CCG/CGG and CGC/GCG were the most prevalent. The most frequent dimer motifs found were GA/TC, CA/TG and AT/TA. Annotation with Gene ontology (GO) confirmed that the highest proportion of ESTs had functions related to transferase activity followed by protein binding. Twenty one primer pairs from 11 WRKY family transcription factors and 10 disease resistance proteins (DRP) were used to survey the polymorphism in a set of 24sugarcane clones. High genetic diversity was observed for the loci WRKY2, WRKY 6, WRKY 5, WRKY8, WRKY9b, WRKY11, DRP1, DRP5 and DRP10 among the cultivated species clones. Erosion of the alleles WRKY1a, WRKY4, WRKY 9a, WRKY10, DRP6 and DRP 9b among the cultivated species group is an important observation in the present study. Neutrality tests revealed a deficiency of heterozygotes for the alleles WRKY 9, WRKY 11, DRP 7 and DRP 10 indicating a pattern of positive selection among the cultivated species clones. The new genic SSR markers specific to WRKY and other stress factors identified from this study would facilitate the QTL identification and marker-assisted selection due to its association with the functional regions of the sugarcane genome.

Key words: Sugarcane, Genic SSRs, WRKY, stress transcription factors, AMOVA, diversity index

Introduction

Breeding of improved cultivars of sugarcane is difficult because of the complexity of the genome and the long duration required for the identification of a commercial variety. Application of molecular markers eases any traditional breeding approach. Role of molecular markers in sugarcane breeding is of high value because of lack of whole genome sequence and transcriptome assembly in order to utilize them for improvement. In the last two decades, SSRs have often been exploited for sugarcane improvement (Cordeiro et al. 2001;Aitken et al. 2005; Edme et al. 2006; Pan 2006; Hameed et al. 2012). Brazilian workers have developed large number of ESTs from various development stages for sugarcane (Sugarcane EST Project-SUCEST) (Vettore et al. 2001) and the database was further

R.M. Shanthi*, Ravinder Kumar, Division of Crop Improvement, Sugarcane Breeding Institute, Coimbatore-641007
Padma Nimmakayala and O.U.K. Reddy, Department of Biology, West Virginia State University, USA
*email:rmshanthi288_@rediff.com

surveyed to find out EST-SSRs (Pinto et al. 2004). A large number of EST-SSRs are available for different traits in sugarcane (Parida et al. 2010;Marconi et al. 2011; Singh et al. 2013). Most of these EST-SSRs represent genes associated with important metabolic processes such as photosynthesis, carbohydrate metabolism, sugar transport, amino acid metabolism and biotic and abiotic stress response mechanisms. Despite the availability of a large number of SSR markers for sugarcane improvement, genic-based EST-SSRs for improvement of stress tolerance traits are inadequate.

A number of transcription factors and *cis*-regulatory sequences in plants have been identified to have a significant role in abiotic and biotic stresses. Among them, the plant WRKY transcription factors, comprising a large family of regulatory proteins, play an important role in response to various stresses (Pandey and Somssich 2009). In rice, different WRKY members showed enhanced tolerance to heat (Wu et al. 2009), cold tolerance (Yokotani et al. 2013) and plant-pathogen interaction (Tao et al. 2009). Information on the allelic diversity in WRKY and other stress transcription factors is limited in sugarcane. A study was carried out to characterize the frequency and relative distribution of transcript repeat motifs in WRKY and other stress factors in the coding regions; design primers flanking the repeat-motifs; and study their efficiency in the assessment of molecular diversity among a set of wild and cultivated species clones in sugarcane.

Materials and methods

A microsatellite search was conducted on a high performance computer using a version developed with the Perl script available as Simple Sequence Repeat Identification Tool (SSRIT) at Cotton Microsatellite Database (CMD) (http://www.cottonssr.org). For development of genic SSR markers, unigene data sets such as the gene indices from the Institute of Genomic Research (TIGR: http:/

/www.tigr.org/tdb/tgi) were used. This program was operated with the parameters set to detect di- to hexanucleotides of specified lengths. To further examine the location of SSRs in the sequences in relation to the putative coding region, the SSR server used the FLIP program (Brossard, 1997) which is available through the Organelle Genome Megasequencing Project (OGMP). Using the FLIP output, the longest ORF is identified and the relative SSR location is reported. A batch of 157190 sequences was uploaded in FASTA format using the CMD SSR tool. On completion of the processing, we could download 1) a summary report of the SSR analysis, 2) a library file of the uploaded sequences, 3) a library file of the SSR containing sequences and 4) an excel file of the individual properties of the SSR-containing clones. The sequence properties included sequence name, length of the SSR-containing sequence, repeat(s) motif and number, SSR start/stop position, ORF start/stop position, primer pairs, SSR location relative to the ORF, and GC content of the sequence. The sugarcane clones used for the polymorphism survey in this study comprised11 wild, 10 cultivated species clones and three commercial hybrids (Table 1).

Primers were designed using Primer3 (Rozen and Skaletsky, 1999) with the following specifications: optimum primer length of 20 nucleotides (nt) (18-26 nt), optimum melting temperature of 50°C (45-55°C), an optimum product size of 125 base-pairs (100-350 bp) and an optimum G/C content of 50% (45-55%). SSR results were run through a Gene Ontology (GO) assignment database in order to assess associations between SSR loci and biological processes, cellular components and molecular function of known genes. A FASTA file with all ESTwas subjected to Blast2GO software and ran against the GO annotated sequences, and the obtained hits were compiled.

Approximately 100 mg of bulked leaf tissue from four grown-up plants per genotype or line was ground

Table 1. Sugarcane clones used for the study

S.No.	Clone	Species / Group
1	28 NG 224	*S. officinarum*
2	57 NG 110	*S. officinarum*
3	H 52	*S. officinarum*
4	Penang	*S. officinarum*
5	Reha	*S. barberi*
6	Saretha	*S. barberi*
7	Katha	*S. barberi*
8	Uba white	*S. sinense*
9	IKRI	*S. sinense*
10	DhaurCalig	*S. sinense*
11	CoC 671	Commercial hybrid
12	Co 740	Commercial hybrid
13	B54142	Commercial hybrid
14	NG 77-73	*S. robustum*
15	IJ 76-545	*S. robustum*
16	NG 77-28	*S. robustum*
17	S. edule	IJ 76-551
18	Iritty-2	*S. spontaneum*
19	SES 194A	*S. spontanuem*
20	Coimbatore local	*S. spontanuem*
21	IK 76-76	*E. arundinaceus*
22	IK 76-99	*E. arundinaceus*
23	M 75-062	*Miscanthus*
24	US 56-0022	*Miscanthus*

to a fine powder using TissueLyser II (Qiagen, Valencia, CA). DNA was then extracted using DNeasy plant mini kit (Qiagen, Valencia, CA) and concentration was quantified with Nanodrop-1000 (Nanodrop technologies, Wilmington, DE).

PCR reaction conditions were as follows: 200 ng of genomic DNA, 0.20 ìM of mixed forward and reverse primers, 1X Buffer (10 mM de Tris-HCl pH 8.2, 50 mM KCl, Triton 0.1%, BSA 1mg/ml), 1.5 mM $MgCl_2$, 0.2 mMd NTPs and 1 U *Taq* polymerase in 10 ìL reaction volumes. Amplification was performed in a GeneAmp PCR 9700 System thermal cycler (Applied Biosystems Inc.) programmed as follows: 94°C for 2 min followed by 35 cycles of 94°C for 30 s, 50-65°C for 30 s,

72°C for 1 min, and a final extension step at 72°C for 10 min. Amplified products were separated on 3% SFR agarose gels. In scoring, stutters were avoided and discernible bands were scored as alleles. Allele sizes were estimated and mapping gels based on comparison with 50 bp molecular weight ladder that were distributed twice on each SFR grade agarose gel. PCR products were sequenced using the procedure proposed by Sanger et al. (1977).

Measures of population genetic diversity viz., Shannon's diversity index (I), gene diversity / heterozygosity (*H*) (Nei, 1973) were obtained using the program *popgene* version 1.31 (Yeh et al. 1997). Population structure and inbreeding were examined by the analysis of variance procedure (Weir and Cockerham, 1984), which provides estimators of Wright's *F*-statistics F_{IS}, F_{IT}, F_{ST} (Wright 1978). These were calculated over all samples, then separately for each population. Linkage disequilibrium (LD) analysis and building Neighbor joining phylogenetic tree was carried out using TASSEL ver. 2.1 (www.maizegenetics.net).

Results and discussion

Development of EST-SSRs through data mining has become fast, efficient and relatively inexpensive compared with that of genomic SSRs. These markers are located at specific regions in the genome, can be detected with high reproducibility, and are multi-allelic, co-dominant, analytically simple and readily transferable (Gupta and Varshney 2000). In this study, a set of 24 sugarcane clones was assayed for the presence of SSRs in the transcribed portion of the genome. Novel microsatellites were mined from the available EST resources; the abundance of SSRs has been explored with respect to WRKY stress related pathways. From data mining, a total of 18275 SSR makers were analyzed for gene prediction/Gene Ontology and for functional gene identification using the sequence similarity search program-BLAST with the tblastx option (the threshold E-value cutoff at 1e-6 against the Swiss-

Table 2. Information on the microsatellite specific to WRKY and disease resistance related proteins

	Primer	SSR allele	Seq. length (bp)	Motifs	No. of repeats	Location of SSR	Start	End	Forward primer sequences (5'-3')	Reverse Primer sequences (5'-3')
1	TA30836_4547	*WRKY1	985	GT	10	3'UTR	656	675	TTGAATTTTCGAGCCCAAAC	CAAGGAAAGATTGTAGCCGC
2	TA23294_4547	WRKY2	759	GCC	5	ORF	489	503	AAGAGCGAGAGCATGGACAC	TCACCTTTCTGCCGGTACTT
3	CA077905	WRKY3	532	CT	5	5'UTR	20	29	ACCACACCTNCCGAACTGC	TCTTTCTTTGGAGGCAGGAA
4	CA162653	WRKY4	821	CGC	6	ORF	478	495	AAGAGCGAGAGCATGGACAC	AGAGCAACCGNAAGTATGCC
5	CA120996	WRKY5	802	GCG	5	ORF	277	291	CAGAAAGTGGTGAAGGGGAA	CAGCATCTCCAGGGTGTAGG
6	CA109717	WRKY6	635	AAG	8	5'UTR	232	255	TTCCCTAGAGGAAGGGAGGA	CTAGCACAGGATGAACGCAA
7	CA108941	WRKY7	655	CCG	5	ORF	216	230	CATCCAAGAACCCAACCACT	ATATGGCTCTGGCTCTGGCT
8	CA139234	WRKY8	932	GCG	5	ORF	408	422	CAGAAAGTGGTGAAGGGGAA	CAGCATCTCCAGGGTGTAGG
9	CA079510	WRKY9	717	TA	5	ORF	268	277	TCCTGCCTCCAAAGAAAGAA	TCGAATCAAGGAAACGATCC
10	CA088496	WRKY10	663	AG	5	ORF	38	47	CTACTGGGGAAAGCAAAGCA	AGGAGGAGCCGTAACCTAGC
11	CA159375	WRKY11	727	GCA	5	ORF	167	181	GATGATGAGTGACCTCGTCG	AGAGTTGTAGTTGCGGGCAT
12	TA39046_4547	**DRP1	1031	GCT	5	5'UTR	77	91	GGCACACCTCTAGAGACCCA	AAACAGAAACCGGACAGCAC
13	TA39046_4547	DRP2	1031	CGG	6	ORF	331	348	GCTGTCCGGTTTCTGTTTGT	GCGCTCCTTCTCCTCCAT
14	TA39047_4547	DRP3	1095	GCT	8	5'UTR	59	82	ACACCTCGAGAGACCCAGG	ACTCCTCCTCCTCGCTTAGG
15	CA097819	DRP4	585	CG	6	ORF	137	148	CTCCTCCACCTCAAGTCCCT	TGGAGTGGGAGCAGAAGG
16	CA221449	DRP5	960	CG	5	ORF	151	160	GCTGGACAAGTACAGCGACA	TCGAGGCTCTGGTACACCTT
17	TA30918_4547	DRP6	857	GCA	5	ORF	97	111	GTCCGCACATACTCACGGT	GAGGAGGAGGAGGAAGAGGA
18	TA32228_4547	DRR7	1063	TCC	5	ORF	176	190	GCCGGTCCCATACATAACAC	TGCATGAAGAAGCTCAGGTG
19	CA182634	DRP8	672	GCA	6	ORF	55	72	GTTCGCAGTTCGAGGGTC	GTCGTCGATTACCGAGGTGT
20	CA185225	DRP9	929	TCC	5	ORF	101	115	AACCACGCCCATTCCTTC	TGCATGAAGAAGCTCAGGTG
21	TA39512_4547	DRP10	801	CGA	5	ORF	611	625	ATCTACGACGAGACGAGGGA	AAGGGGATCGGAGAGGTAGA

*WRKY family transcription factor (WRKY), **disease resistance protein (DRP)

Prot database (http://web.expasy.org/docs/swiss-prot_guideline.html). SSRs were located in ORFs and UTRs of major genes of WRKY transcription factors and pathogen disease resistance responsive protein among the wild and cultivated species clones (Table 2). Higher number of repeat units (13586) was found in open reading frames rather than those numbers (4689) found in 5' and 3' UTRs (Fig 1). Dimer repeats located in ORF regions were 4865 in comparison with 1641 and 1723 dimers in 5'UTR and 3'UTR regions respectively**.** In addition, higher number of trimers (8527) were located in ORF regions compared to 4868 (dimer motifs) and 191 tetra repeat units (Fig. 2). Compared to dimers, trimers are predominantly located in the protein encoding regions. These results are in agreement with previous reports that trinucleotide SSRs are the most abundant type in *Arabidopsis* ESTs (Cardle et al. 2000) and in exons of genomic DNA sequences in all eukaryotes studied (Toth et al. 2000).

The frequency of SSRs in non-redundant ESTs more accurately reflects the density of SSRs in the transcribed portion of the genome. Repeat numbers of various dimers, trimers and tetramers and their frequency of occurrence are summarized in Table 3. Dimer motifs were GA/TC (3361), CA/

Fig. 1. Location of transcript SSR motifs in *S. officinarum*
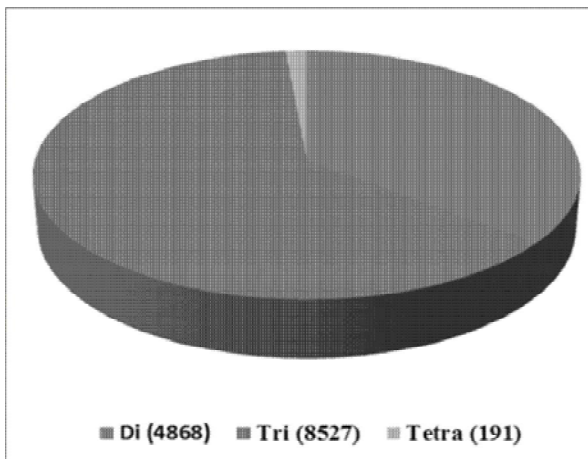
Fig. 2. Distribution of SSR motifs in the ORF region of the genome

Table 3. Distribution of EST- SSRs based on the number of repeat units

Repeat Number	Di	Tri	Tetra	Total
5	5007	6061	238	11306
6	1297	2171	47	3515
7	612	913	15	1540
8	330	327	10	667
9	205	149	4	358
10	118	49	2	169
11	77	29	3	109
12	49	10	2	61
13	61	10	3	74
14	43	3	2	48
15	35	4	2	41
16	46	3	0	49
17	36	2	2	40
18	29	1	0	30
19	18	1	0	19
20	28	0	1	29
21	19	1	0	20
22	12	2	0	14
23	12	1	0	13
24	18	0	0	18
25	12	0	1	13
26	10	0	0	10
27	16	1	0	17
28	11	1	0	12
29	6	0	0	6
30	8	0	0	8
Others	88	1	0	89
Total	8203	9740	332	18275

TG (1696), TA/TA (1585) and CG/GC (1561) totalling 8203, which is 45% of total 18,275 repeats motifs. The predominant trimer motifs found were GCC/GGC (8.95%), CGC/GCG (8.69%) and CCG/CGG(8.59%), AGC/GCT(3%), GCA/TGC(2.6%), CAG/CTG(2.4%), CTC/GAG(2%), AGG/CCT(2.0%), GGA/TCC(1.8%), GAC/GTC(1.5%), CGA/TCG(1.5%) with 1636, 1588, 1570, 546, 467, 436, 358, 356, 326, 268 and 266 occurrences in 9740 trimers. Rest of the trimers that were considered as minor was 1923. Overall, tri-nucleotide repeats dominate the population and five-time repeats of di-, tri- and tetra- were the most prevalent in the SSR population. In rice, 60% of EST-derived microsatellite sequences were represented by the motifs CCG, ACG, AGG and ACC (Temnykh et al. 2000), whereas in maize CCG/GGC and AGG/CCT were most abundant (Chin et al. 1996). Earlier observations in the expressed sequence tags and unigene sequences of cereal genomes revealed the relative abundance of GC rich trinucleotide repeat motifs (Varshney et al. 2005). Similar studies in sugarcane indicated CCG as the most common motif (Cordeiro et al. 2000) that supports the finding in the present study.

Annotation of the EST-SSRs revealed the polymorphisms within the transcripts and that the main functional category was related to transcription and post-transcriptional regulation. From the annotation of the sugarcane EST-SSRs to other plant genome databases, the distribution of the best Blast hits is presented in Fig. 3. Out of 18,275 SSR containing sequences, 45.93% blast hits were identified as similar to *Oryza sativa* and 36.47% were not identified. Rest of them was matched to several grass genomes as well as to *Arabidopsis.* Of the 18,675 SSR loci found, 10,205 had gene ontology assignments. Gene ontologies pertaining to molecular function assignment level revealed a predominance of transferase activity (15.04%) and protein binding (12.72%), suggesting that these are representatively higher in genome. The higher occurrence of ontology levels to the functions described indicated potentiality of using these microsatellites as markers to saturate associated pathways. The number of SSR-ESTs that produced no hit was 42 (10.85%). This indicates presence of sequences encoding proteins which are specific to sugarcane stress factors or proteins are present in other plant/animal systems but are still not reported.
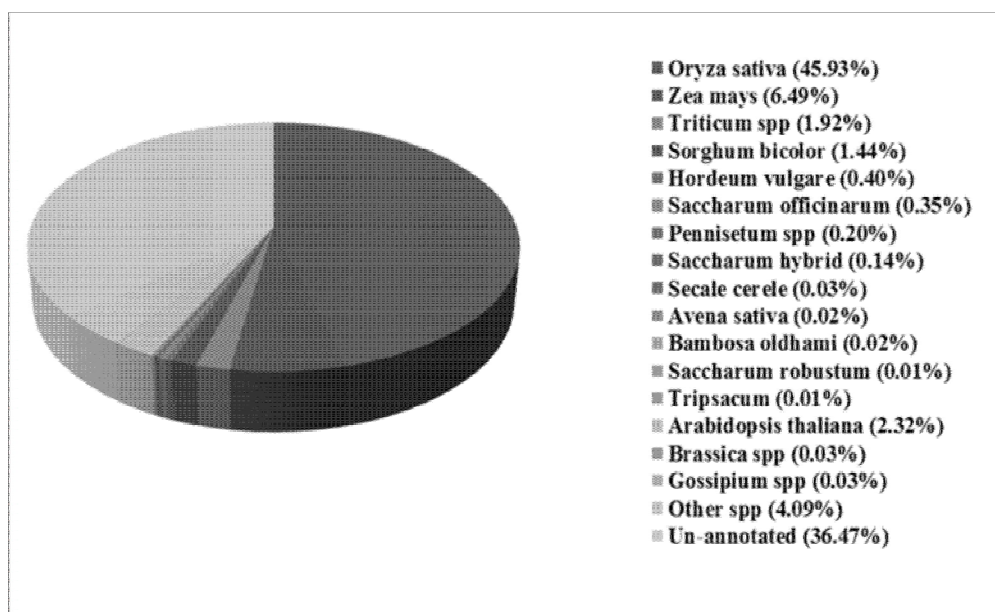

Fig. 3. Functional annotation of SSR motifs containing *S. officinarum* sequences with other genomes

Role of WRKY transcription factors in plant abiotic stresses has been demonstrated and many WRKY genes behave strongly and rapidly induce expression in response to certain abiotic stresses, such as wounding, drought or salinity, indicating their regulatory function in these signalling pathways. Twenty one primer pairs were designed to flanking SSR motifs from 11 WRKY family transcription factors and 10 disease resistance proteins (DRP) and used for amplifying a set of wild species and cultivated species clones. Information on the WRKY specific alleles amplified and their product sizes are presented in Table 4. Analysis of molecular variance (AMOVA) can be used to describe partitioning of the total genotypic variation among and within the groups. In this study, AMOVA based on SSR allelic frequencies showed that the two groups (wild and cultivated species) as well as the taxa within the groups were significantly different (Table 5). Gene diversities (Shannon's Index and Nei's heterozygosity) among the wild species group were high for the loci WRKY 1b, WRKY 3, WRKY 7, DRP2, DRP 3, DRP 4, DRP 7, DRP 8 and DRP 9a when compared to the other (cultivated species) group. Among the cultivated species group, gene diversities were high for the loci WRKY2, WRKY 6, WRKY 5, WRKY8 WRKY9b, WRKY11, DRP1, DRP5 and DRP10 when compared to the wild species group. Another significant finding in this study is the erosion of alleles WRKY1a, WRKY4, WRKY 9a, WRKY10, DRP6 and DRP 9b among the cultivated species group (Fig. 4). Positive values of Wright's fixation index (F_{IS}) as a measure of heterozygote deficiency were observed for WRKY1b (0.60), WRKY4 (1.00), WRKY5 (0.73), WRKY7 (1.00), WRKY8 (0.68), DRP1 (0.375), DRP3 (0.70) and DRP10 (0.35) in domesticated group. Loci wise significant F_{IS} and F_{ST} values were presented in Fig. 5. Evens-Watterson test of neutrality was performed to cultivated species group to resolve the loci that are under positive and diversifying selection (Fig. 6). WRKY2, WRKY3, WRKY6, WRKY10 and DRP1

Table 4. Microsatellite alleles amplified and allele size for WRKY and other stress transcription factors

Primer	SSR allele	Allele Size(bp)
TA30836_4547	WRKY1	220
		225
		235,247
TA23294_4547	WRKY2	245
		267
CA077905	WRKY3	257
CA162653	WRKY4	240
		295
CA120996	WRKY5	260,282
CA109717	WRKY6	231,248
CA108941	WRKY7	215
		232,240
CA139234	WRKY8	263,290
CA079510	WRKY9	107
		130
		119,138
CA088496	WRKY10	280
		287
CA159375	WRKY11	210
		228
		235
TA39046_4547	DRP1	193,200
TA39046_4547	DRP2	143,150
		163
		185
TA39047_4547	DRP3	133,162
		148
CA097819	DRP4	107
		118,131
CA221449	DRP5	120
		132
TA30918_4547	DRP6	112
		125
TA32228_4547	DRP7	255
		272
		168
CA182634	DRP8	242
		260
CA185225	DRP9	168
		175,186
TA39512_4547	DRP10	231
		243

WRKY family transcription factor (WRKY), disease esistance protein (DRP)

Table 5. Analysis of molecular variance among populations and within populations

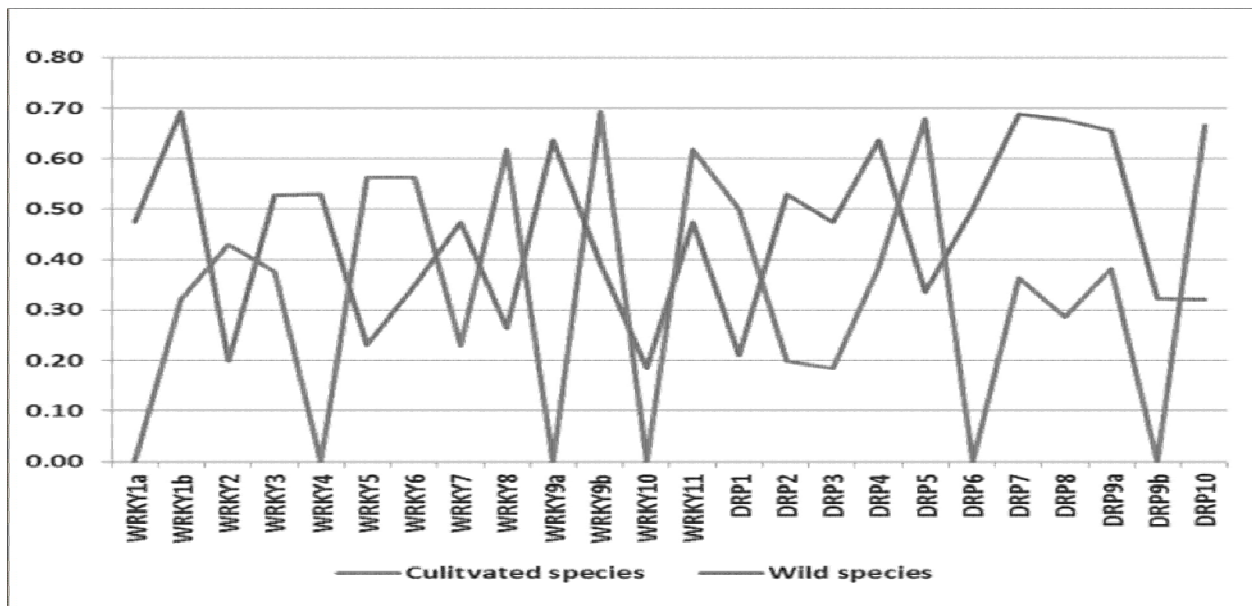| Source of variation	df	Sum of squares	Variance components	Percentage variation
Among populations	1	8.320	0.190	3.91**
Within populations	48	182.006	4.660	96.08**
Total	49	190.326	4.853	

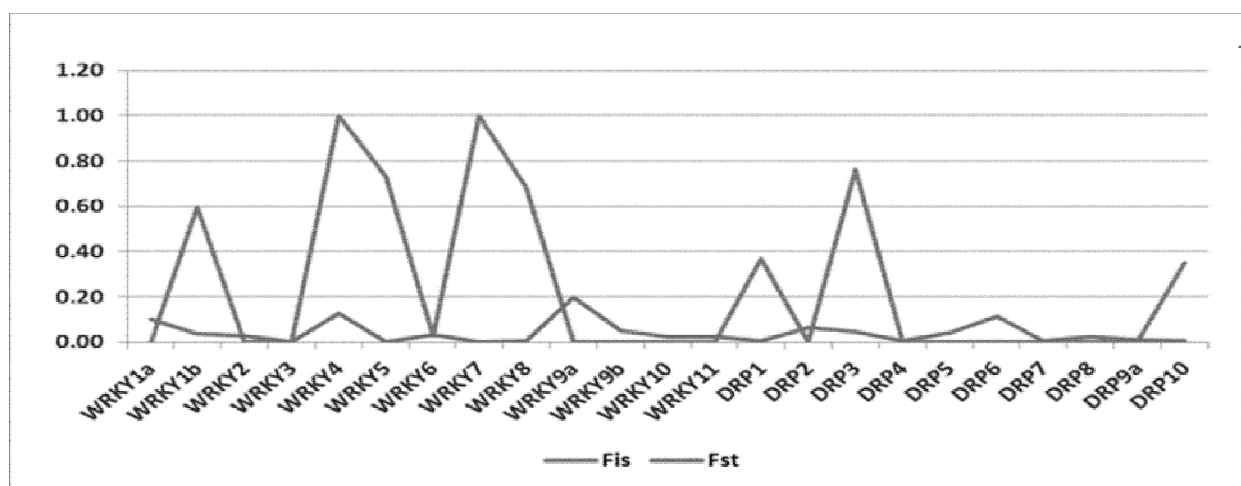**Fig. 4. Estimates of gene diversity for SSR markers related to WRKY and other disease resistance protein among the wild and cultivated species clones**

Fig. 5. Wright's fixation indices for WRKY and stress factor SSR alleles among two different (wild and cultivated) populations in sugarcane
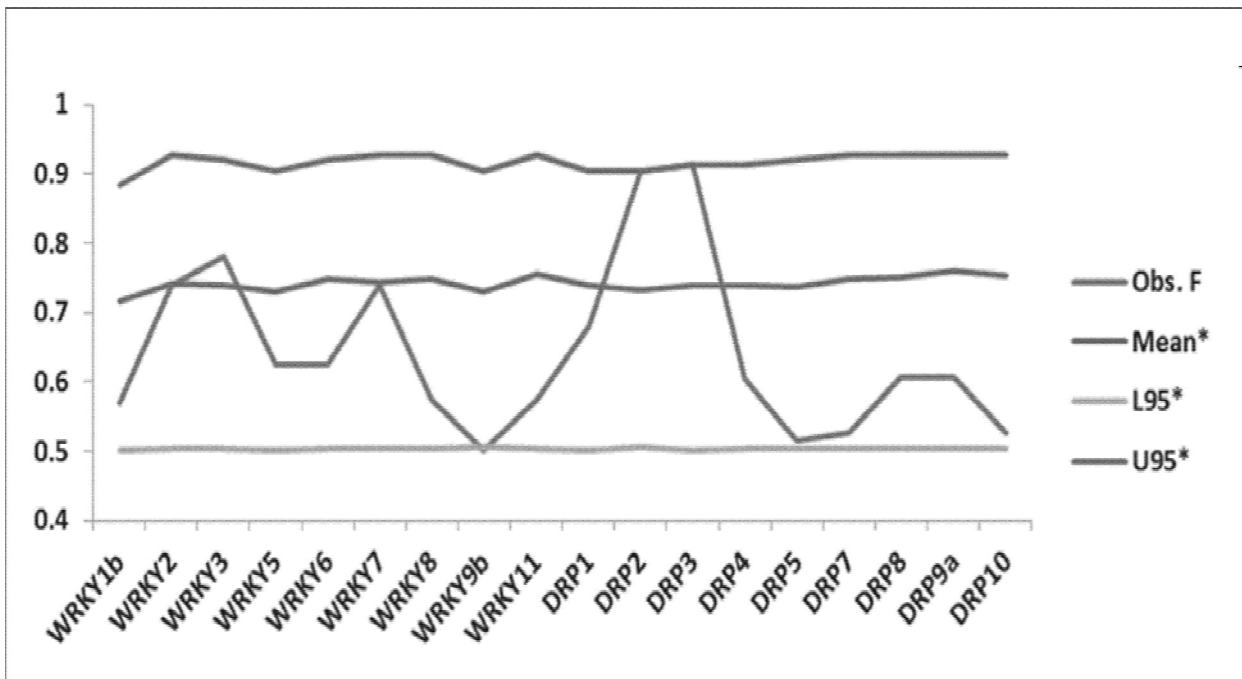
Fig. 6. Ewens-Watterson neutrality test for WRKY and disease resistance protein related SSR markers for cultivated species clones

from cultivated group exhibited an excess in genetic diversity relative to the number of alleles indicating diversifying selection. WRKY1b, WRKY8, WRKY9b, WRKY11, DRP7 and DRP 10 were significantly below the mean curve with substantial heterozygote deficiency, a pattern of positive selection. For Linkage disequilibrium (LD) between two multi-allelic loci, r^2 (statistical coefficient of determination) is the widely used measure of LD for each pair of alleles, or even for overall LD between all the alleles at two loci (Gupta etal.2005). D' (standardised LD measure) is informative for comparison of different allele frequencies across loci and measures only recombination differences, whereas r^2 summarizes both recombination and mutation events. Also, r^2 is indicative of how markers might be correlated with the QTL of interest, so r^2 is often preferred for association studies (Abdallah et al. 2003). Pairwise LD is depicted as a color-code triangle plot (Fig.7) based on significant pairwise LD level (r^2, p-value as well as D') that helps to visualize the blockof loci in

significant LD. Significant evidence for LD was noted between WRKY4 and WRKY1a (r^2 =1.00 and P=0.004), DRP3 and WRKY1a (r^2 =1.00 and
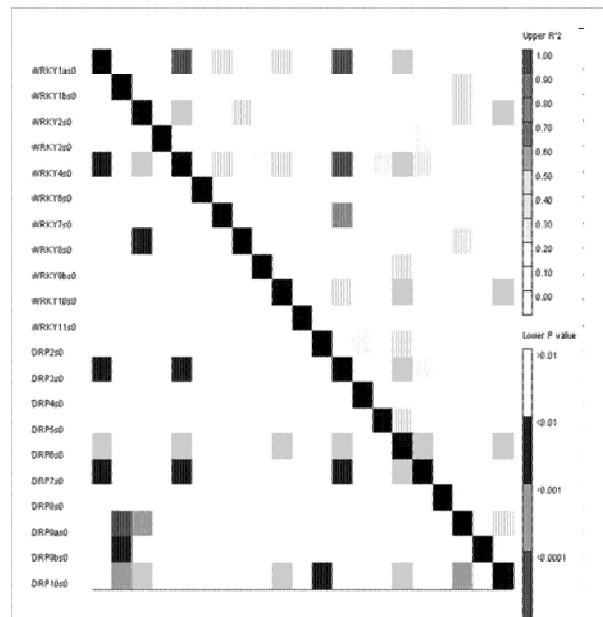

Fig. 7. The TASSEL generated triangle plot for pairwise LD between marker sites for WRKY and other stress factors

P=0.005), DRP3 and WRKY4 (r^2 =1 and P=0.006), and DRP3 and WRKY7 (r^2 =0.63 and P=0.01). In the sets of wild and cultivated sugarcane clones examined in this study, LD as r^2 is present on a scale that could be useful for association mapping. Neighbour joining analysis was performed to build a phylogenetic tree (Fig. 8), which clustered wild and cultivated species clones in two separate clusters with a minor exception of clustering of a *S. spontaneum* (SES194A) with cultivated group.

Thus, genic SSR markers specific to important agronomic traits seem to be very useful for sugarcane germplasm accessions to perform association mapping analysis. Further it is interesting to note that they are clearly conserved and highly transferable across the *Saccharum* complex. This analysis provided information on evolution of various allele specific markers pertaining to WRKY and other stresstranscription factors that resolved patterns of selection and provided important clues to domestication among a set of wild and cultivated species clones in sugarcane. The identification of SSRs in transcripts encoding proteins involved in transcriptional regulation and other functions from the current study provides pertinent markers for applications such as mapping, molecular breeding and QTL analysis in sugarcane.

Acknowledgements

References

Abdallah J, Goffinet B, Cierco-Ayrolles C, Perez-Enciso M (2003) Linkage disequilibrium fine mapping of quantitative trait loci: A simulation study. Genetics Selection Evolution 35: 513 - 532.

Aitken KS, Jackson PA, McIntyre CL (2005) A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. Theor Appl Genet 110, 789-801.

Brossard N (1997) *FLIP: a Unix Program used to find/translate orfs.* bionet software. http://www.bch.umontreal.ca/ogmp/manlinks/flip.txt

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

Chin ECL, Senior ML, Shu h, Smith JSC (1996) Maize simple sequence repetitive DNA sequences: abundance and allele variation. Genome 39: 866-873.

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to Erianthus and Sorghum. Plant Sci1 60: 1115-1123.

Cordeiro GM, Taylor GO, Henry RJ (2000) Characterisation of microsatellite markers from sugarcane (*Saccharum* sp.), a highly polyploid species. Plant Science 155: 161-168.

Edme SJ, Glynn NG, Comstock JC (2006) Genetic segregation of microsatellite markers in *Saccharum officinarum* and *S. spontaneum.* Heredity (Edinb) 9**7**: 366-375.

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

Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol Bio l57: 461-485.

Hameed U, Pan YB, Muhammad K, Afghan S, Iqbal J (2012) Use of simple sequence repeat markers for DNA fingerprinting and diversity analysis of sugarcane (Saccharum spp) cultivars resistant and susceptible to red rot. Genet Mol Res 11: 1195-1204.

Marconi TG, Costa EA, Miranda HR, *et al.* (2011) Functional markers for gene mapping and genetic diversity studies in sugarcane. BMC Res Notes 4: 264.

Nei M (1973) Analysis of Gene Diversity in Subdivided Populations. Proceedings of the National Academy of Sciences 70: 3321-3323.

Pan YB (2006) Highly polymorphic microsatellite DNA markers for sugarcane germplasm evaluation and variety identity testing. Sugar Tech 8: 246-256.

Pandey SP, Somssich IE (2009) The role of WRKY transcription factors in plant immunity. Plant Physiol 150: 1648-1655.

Parida SK, Pandit A, Gaikwad K, Gaikwad K Sharma TR, Srivastava PS, Singh NK, Mohapatra T (2010) Functionally relevant microsatellites in sugarcane unigenes. BMC Plant Biol 10: 251.

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

Rozen S, Skaletsky H (1999) Primer3 on the WWW for General Users and for Biologist Programmers In: T Bioinformatics Methods and Protocols, pp. 365-386.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences 74: 5463-5467.

Singh RK, Jena SN, Khan S, Yadav S, Banerjee N, Raghuvanshi S, Bhardwaj V, Dattamajumdar S K, Kapur R, Solomon S, Swapna M, Srivastava S, Tyagi AK (2013) Development, cross-species/genera transferability of novel EST-SSR markers and their utility in revealing population structure and genetic diversity in sugarcane. Gene 524: 309-329.

Tao Z, Liu H, Qiu D,Zhou Y, Li X, Xu C, Wang S (2009) A pair of allelic WRKY genes play opposite roles in rice-bacteria interactions. Plant Physiol 151: 936-948.

Temnykh S, Declerck G, Lukashova A, Lipovich L, Cartinhour S, Mc Couch S (2001) Computational and experimental analysis of microsatellites in rice (Oryza sativa L.): frequency, length variation, transposon association and gentic marker potential. Genome Res 11: 1441-1452.

Toth G, Gaspari Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res 10: 967-81

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

Vettore AL, Silva FRd, Kemper EL, Arruda P (2001) The libraries that made SUCEST. Genetics and Molecular Biology 24: 1-7.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. Evolution 38: 1358-1370.

Wright S (1978) Variability Within and Among Natural Populations, University of Chicago Press, Chicago.

Wu X, Shiroto Y, Kishitani S, Ito Y, Toriyama K (2009) Enhanced heat and drought tolerance in transgenic rice seedlings overexpressing OsWRKY11 under the control of HSP101 promoter. Plant Cell Reports 28: 21-30.

Yeh FC, Yang R-C, Boyle (1997) POPGENE, the user-friendly shareware for population genetic analysis., University of Alberta.

Yokotani N, Sato Y, Tanabe S (2013) OsWRKY76 is a rice transcriptional repressor playing opposite roles in blast disease resistance and cold stress tolerance. J Exp Bot, 10.1093/jxb/ert1298.