

## Bioinformatic Prediction of SNP Markers in WRKY Sequences of Palms

Manju, K. P<sup>1</sup> and Arunachalam, V<sup>2,3\*</sup>

### Abstract

WRKY transcription factors are unique to plants and performs many imperative functions mainly disease resistance. In the present study we have analyzed the WRKY transcription factor gene sequences to assess the variation at single nucleotide level. We have retrieved 525 sequences of WRKY genes of palms of 334 Kb size. The sequences were purified by employing EST trimmer and were clustered into 31 contigs using CAP3. Single nucleotide Polymorphisms (SNP) and insertion/deletions (indels) were detected in contigs using the AUTOSNP software. Alternatively candidate SNP containing contigs were aligned by Clustal X to locate the SNPs. Results from these two methods were compared and false SNPs were eliminated. Finally, about 568 SNPs were found including 250 transitions, 120 transversions and 198 indels. The SNPs were seen at a frequency of 2.84/100bp in the WRKY sequences of palms. Primers were designed flanking to SNP/ indel sites with potential as markers in palms. We could obtain two novel WRKY-SNP markers (WRKY 7 and WRKY 12) which are not reported before in palms.

**Key words:** SNP, Arecaceae, *Cocos nucifera*, WRKY

---

<sup>1</sup>DIT Agribioinformatics Promotion Centre, Central Plantation Crops Research Institute, Kudlu, Kasaragod-671124, Kerala, India.

<sup>2</sup>Molecular Biology Lab, Central Plantation Crops Research Institute, Kudlu, Kasaragod 671124, Kerala, India.

<sup>3\*</sup> Present addresses: Horticulture Section, ICAR Research Complex for Goa, Old Goa North Goa Goa 403 402 India \*corresponding author's email [v.arundevi@gmail.com](mailto:v.arundevi@gmail.com)

## Introduction

WRKY transcription factors are a superfamily of proteins containing one or two highly conserved WRKY domain among the family members. WRKY domains have conserved regions constituted of 60 amino acid sequence with a WRKYGQK at its N-terminal end, together with a C<sub>2</sub>H<sub>2</sub> or C<sub>2</sub>HC novel zinc-finger like motif. This domain often exhibit sequence-specific DNA binding activity and have potential for activating or repressing transcription of target genes. This domain binds specifically to the DNA sequence motif (T) (T) TGAC(C/T), which is known as the W box. The invariant TGAC core is essential for function and WRKY binding (Eulgem *et al.*, 2000). Thus WRKY proteins play significant roles in development and response to biotic and abiotic stresses. Thus they are very important regulators of defense transcriptome and disease resistance (Eulgem *et al.*, 2000). Studies indicated that WRKY genes play a role in the signaling cascade of innate immunity in many other plant species such as Arabidopsis, tobacco, rice and parsley (Wu *et al.*, 2005). Some WRKY TFs are manipulated by pathogen effectors to promote virulence. Multiple pathogen effectors are targeted to host nuclei and modify expression of the defense transcriptome. Also the WRKY transcriptional networks provide a suitable environment to respond and arrest pathogens, simultaneously by preventing the defense responses that are deleterious to plants itself (Pandey *et al.*, 2009).

Evidences suggest that the WRKY gene families arose during evolution through duplication. The phylogenetic analysis of WRKY genes can serve as a useful guide for studying their roles in plants and the phylogenetic analyses of the WRKY domain sequences provide support for the hypothesis that gene duplication of single- and two-domain WRKY genes, and loss of the WRKY domain, occurred in the evolutionary history of this gene family in rice (Xie *et al.*, 2004). Also the WRKY transcriptional network may provide the proper balance to respond quickly and efficiently to deter pathogens, but at the same time to restrict

defense responses that can be detrimental for plant growth and development (Lopez *et al.*, 2007). Diversity in coconut populations was analyzed using SNP and SSR markers derived from WRKY gene families (Mauro-Herrera *et al.*, 2006). WRKY gene based marker system is a rapid and simple method for generating sequence specific markers for plant gene families (Kim *et al.*, 2007).

Molecular markers provide a link between genotype and phenotype, for the production of molecular genetic maps and to assess the genetic diversity within and between related species. An important marker, Single Nucleotide Polymorphisms (SNPs) are single DNA sequence variation in the genome, which usually bi-allelic in nature, so can be easily assayed. SNPs are becoming the marker of choice for molecular genetic analysis. SNPs represent sites, where the DNA sequences differ by a single base. SNPs represent the most common form of genetic variation in both plants and animals, and play a key role in revealing the molecular mechanisms underlying traits. SNPs are increasingly becoming the marker of choice in genetic analysis and are used routinely as markers in agricultural breeding programs (Gupta *et al.*, 2001). They also have many uses in human genetics, such as for the detection of alleles associated with genetic diseases and the identification of individuals (Nikiforov *et al.*, 1994). SNPs are invaluable as a tool for genome mapping offering the potential for generating very high density genetic maps, which can be used to develop haplotyping systems for genes or regions of interest. The low mutation rates of SNPs also makes them excellent markers for studying complex genetic traits and as a tool for understanding of genome evolution. None of the other markers is as frequent as SNPs which was discovered later. Variation in the DNA sequence can affect how organisms develop diseases and respond to pathogens, chemicals, drugs, vaccines and agents. But as far as the reports are concerned, the discovery and characterization of SNPs are highly expensive and toil. While, the electronic mining of SNPs from the sequence sets furnish the cheapest supplier of plentiful polymorphic marker (Gu *et al.*, 1998; Taillon-

Miller et al., 1998; Buetow et al., 1999; Picoult-Newberg et al., 1999).

Recently Meerow *et al.*, (2009) reported gene sequences of seven WRKY families in palms. This study reports a large WRKY gene sequences in palm which including the earlier known WRKY gene sequences formed the material for our study.

The WRKY conserved domains were useful in the development of molecular markers. Through this study we have analyzed 7 different WRKY loci and located SNP markers within the loci using two software tools. This study is helpful to assess the genetic diversity of palm species using WRKY gene derived SNP markers.

## Materials and Methods

### Sequence Retrieval

We have retrieved 525 sequences of WRKY genes of (WRKY2, WRKY6, WRKY7, WRKY12, WRKY16, WRKY19, and WRKY21) palms (73 species) of 334 Kb size from the nucleotide public domain NCBI (<http://www.ncbi.nlm.nih.gov/>). Sequences were retrieved as FASTA format for subsequent analysis to explore the frequency of SNPs in WRKY genes in palms.

### Sequence Pre-processing

The interpretation of any analysis performed on a contaminated sequence can be perplexed by the presence of segments of foreign origin and will generate misleading data. This involves vector segment contaminations, ambiguous sequences, distal oligoN series and poly A (T) tails. Marker mining was performed only after the driving out of these erroneous regions. In the present study, this was accomplished by EST trimmer software (Thiel, 2001), a perl script useful for pre-preprocessing the sequences. All repeating nucleotides of set type "T" from 5' end and "A" from the 3' end were purposefully removed. Also non sequenced regions results from the potentially accumulating sequencing failures containing a minimum of X/N repeats were also removed.

### Sequence Assembly

CAP3 (Huang *et al.*, 1999), a perl program embedded with Auto\_snip, was used to generate clusters and generate contigs from the given FASTA formatted sequence of WRKY. It had given the -ace format file as output along with corresponding files such as contigs, singlets, info and quality files. The contigs clustered in the -ace file were being used by the onward program to mine the polymorphic regions.

### SNP Detection

Once a proper substrate data set is attained, then later will be the matter of ascertain true allelic variation from the sequencing errors (Marth *et al.*, 1999). This issue could be remedied using a perl script Auto\_snip/d2SNP (version 1.0) to detect SNPs and insertion/deletions (indels). This script takes a fasta or ace format file, aligns the sequences and detects SNPs within the alignment. Thus from the assembled contigs, Auto\_snip program generated an HTML format output file to allow the user to browse through the SNP results. A list of SNP sites, contig informations, positions of short insertions/deletions could extract.

### Multiple Sequence Alignment

In order to scrutinize the result, SNP containing contigs were subjected to alignment by clustal X. This step attempts to discern between true polymorphism and false positives. Clustal is windows interface software for the multiple sequence alignment of nucleic acid and protein sequences. It provides an integrated environment for performing multiple sequence alignments and analyzing the results and also to locate single base substitutions and indels.

### Primer Designing and Verification

On the basis of the predicted SNPs, primer pairs were designed keeping the focus on the marker regions in the specific polymorphism containing sequences. Primers were designed using the softwares such as FastPCR / Primer3. Primers were derived from the upstream and downstream parts of the conserved sequences of WRKY. Later on the primer dimers/self annealing and loopings were healed and the

resultant primers were verified using the stand alone Oligoanalyzer software. Oligo analyzer is a simple tool to evaluate the physical properties of the primer, like Tm, GC%, primer loops, primer dimers and primer-primer compatibility.

### Results and Discussions

Available 525 sequences of WRKY genes from 73 species in 16 genera of *Arecaceae* family were explored in the present study. The pre-processing and clustering had made the sequences unstained and free from redundancy. The duplicated sequences were assembled into 66 contigs keeping back the 54 singletons. From the 66 contigs generated by CAP3 from multiple individuals sharing the common traits, only contigs with four or more sequence reads was undergone SNP verification leading to a total contigs number of 30.

Since the SNP detection was upon single nucleotide changes, the *in silico* SNP prediction will be highly sensitive to errors and less reliability. In fact, a greater number of mismatches were identified at the clustering stage, and many of these mismatches were associated to base-calling or sequencing errors. In this study, from the direct preliminary prediction of SNPs from Auto\_snip program, we could detect a total of about 1241 SNPs, including 374 transitions, 345 transversions and 522 indels from 50 contigs (Table 1). The SNPs were seen at a frequency of 5.8/100bp. Since this type of polymorphism is the change in a single nucleotide, the non-redundant sequences are clustered and aligned to detect the single changes. The principle of multiple sequence alignment by clustal moves on with the progressive alignment and weightage for every match, we have customized the algorithm as to represent the base wise comparison in each column. The polymorphic sites shown were then scanned manually. Thus we could predict WRKY palm SNPs devoid of false positives to an extent.

The presence of a single base polymorphism from a set of sequence data can be hindered by the unending high through put sequencing errors. Rather than the mere

prediction of polymorphic sites using particular programs or tools the analysis and verification of the electronic markers must be performed. False outputs have been used to accumulate, but the filtering of relevant polymorphism among the junk materials will be the key and critical step in these kinds of bioinformatic predictions. Thus in order to obtain relevant SNPs as far as possible, we have screened the datasets twice using two different programs to monitor the hypocritical candidates. Thus we could hamper the erroneous candidates on lend by the multiple sequence alignment program giving significant ones.

So in total 566 polymorphic regions are predicted in palm WRKY sequences with a frequency of 2.84 per 100 bp including the 223 indel polymorphisms (Table 2). A greater number of indel sites were also seen with a frequency of 1.11/100bp kept the remaining 1.72/100bp of SNP polymorphisms. The percentages of A↔G, T↔C, A↔C, T↔G, A↔T and G↔C were 18, 22, 4, 4, 8 and 3 (Table 3). Of the 566 SNPs, 228 were transitions (40%) and 115 were transversions (19%) giving a transition-to-transversion ratio of 2.1:1 (Fig 1). This prediction of ratios found in equivalence with the 2:1 ratio of transitions-to-transversions in mammals (Cheng *et al.*, 2004) and 2.1:1 in Cattles (Lee *et al.*, 2005).

As a general trend we observed that the most frequent type of mutation is that having a base change of either A/G or C/T (Table 3) (Picoult-Newberg *et al.*, 1999). This satisfies our record of more transition rates than transversions/indels with the earlier reports of greater transition frequency by many SNP prediction programs (Garg *et al.*, 1999; Picoult-Newberg *et al.*, 1999;). Again this study is in accordance with the proved fact in monocots that the T↔C transition will be outnumbered than A↔G (Douglas *et al.*, 1998). While the least common type of base change is G↔C. The SNPs predicted in this study can be verified by the PCR techniques and sequencing, so the primers designed keeping the marker region as target using FastPCR/Primer3 tools are listed in the Table 4 and will be useful for the further studies on palm WRKY species.

**Table 1. AutoSNP data: Frequency of nucleotide polymorphisms in WRKY sequences of palms**

WRKY	No. of contigs	Consensus Size (bp)	SNP	Transitions	Transversions	Indels	Frequency (per 100bp)
2	6	4037	625	236	308	81	15.48
6	4	2618	463	161	283	19	17.69
7	4	2990	32	16	9	7	1.07
12	3	2220	45	34	10	1	2.03
16	10	4900	140	37	33	70	2.86
19	12	4760	106	27	25	54	2.23
21	11	9207	332	75	84	173	3.61
	50	30,732	1743	586	752	405	5.67

**Table 2. Predicted frequency of nucleotide polymorphisms/indels in WRKY sequences of palms**

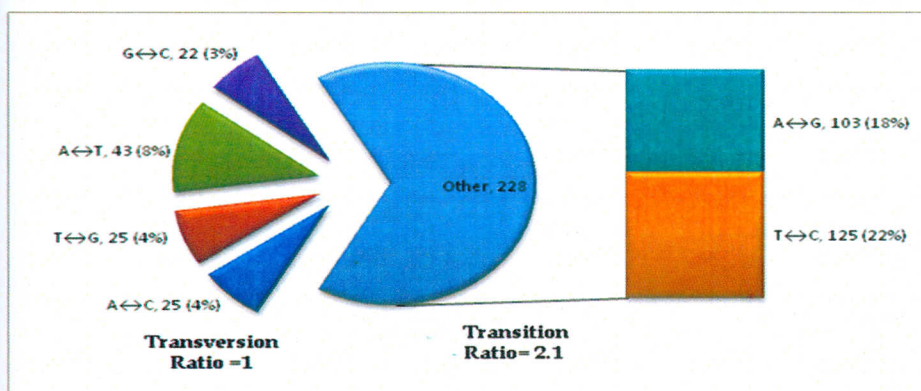
WRKY	No. of contigs	Consensus Size (bp)	SNP	Transitions	Transversions	Indels	Frequency (per 100bp)
2	4	2748	133	72	47	14	4.84
6	2	1519	46	32	4	10	3.03
7	4	2990	32	16	9	7	1.07
12	2	1457	40	31	8	1	2.75
16	6	2949	65	20	9	36	2.20
19	5	2426	63	18	9	36	2.60
21	7	5853	187	39	29	119	3.19
	30	19942	566	228	115	223	2.84

**Table 3. Nucleotide substitution types for the identified WRKY SNPs**

Types	Number of SNPs	Percent	Mutation Type	Ratio*
A↔G T↔C	103 125	18% 22%	Transition	2.1
A↔C T↔G A↔T G↔C	25 25 43 22	4% 4% 8% 3%	Transversion	1

\*Ratio indicates the transition over transversion

**Figure 1. Transition – Transversion ratio in WRKY sequences of palms**



**Table 4a. Designed SNP primers for all the extracted contigs**

Locus	Contig	Forward	Tm (°C)	GC (%)	Reverse	Tm (°C)	GC (%)	Product size (bp)
WRKY2	1	AACTTACCTCGGATTTGGG	51.8	47.4	CACTTCAATGCCGGTTAC	51.3	50	416
	2	ACTCTAAGGTGAACGCTC	51	50	AACTTACCTCGGATTTGGG	51.8	47.4	534
	3	GCGACACATCCACTTCAAT	52.9	47.4	AACTTACCTCGGATTTGGG	51.8	47.4	436
	4	ATGTCACTGGGTGCGAT	53.8	50	AACTTACCTCGGATTTGGG	51.8	47.4	306
WRKY6	1	GTTTCATAGTCTGAGCTTGCC	53.6	47.6	TATCTGGCCCCGACAGCAT	55	55	200
	3	TTCATAGTCTGAGCTTGCC	51.7	47.4	CCAACAAGACCGACATCA	52	50	485
WRKY7	1	ATCCTCAGTTTCGTTTGGTG	53.2	42.9	CCCAACAGAACCAGCTTAG	53.1	52.6	545
	2	AGAAGGTGGTGAAGGCAAT	55.6	47.6	AACAGACCTGGACAACGC	55	55.6	266
	3	CAGAAGGTGGTGAAGGC	53.6	55.6	CAACAGACCTGGACAACG	53.3	55.6	274
	4	TGAACCGATGCAAAGCG	56.1	55.6	AACAGACCTGGACAACGC	55	55.6	511
WRKY12	1	AGATGGGCCCTCGCTTTG	54.9	42.9	TGCCAGAACATACCTTGATT	53.8	58.8	434
	3	CAACCTAGCCGCCGATAC	55	61	ACAAAGCACAAAAGTCCACAT	54	40	693

**Table 4b. Designed SNP primers for all the extracted contigs**

WRKY16	1	AGTTCTCTGGTTTGTGCGG	55.4	50	ACATGACCACAAGGAGCG	54.8	55.6	203
	3	GTCTCAGCTCCTTGTTGGT	53.9	55.6	TACTTCGATCGGGCAAATC	52.1	47.4	261
	5	AGTTCTCTGGTTTGTGCCC	55.3	50	ATAGCGGGTCCCAAATGC	55.2	55.6	366
	6	GTCTCAGCTCCTTGTTGGC	55.6	61.1	AGTGGGAAGGGTACGACA	54.7	55.6	161
	9	TTGTTGCCGGTCTACCC	57	61.1	GATGACCACAAGGAGCCC	55.5	61.1	196
WRKY19	1	GGCTCCAGCGTGTCTCC	58.3	66.7	AGCACACCAAACGAGGC	58.4	61.1	270
	2	TGTAACCCGGCGTAGTCT	52.7	50	GCGGTAGTAATTCCTGCG	53.1	55.6	161
	3	ATATGGCTCCAGCGTGTTC	54.7	52.6	GCGGTAGTAATTCCTGCCA	54.5	52.6	295
	5	TCTTGCTCCCATCATCAT	50.3	44.4	CCAAACCAGGGTAAAGAG	50.2	50	228
	6	CTTCTGCAAACCTAACCACGA	52.7	45	ACGGTAGTAATTCCTGCGA	52.7	47.4	374
	9	CACGCAGAAATATGGCTC	51.1	50	CAATGAGTCTCCACGGT	51.7	50	278
WRKY21	1	CCTTAGCTTGATTTCAGGC	53.2	52.6	TCGTCTGTTGAAGGCGTC	54.5	55.6	646
	4	CCTTAGCTTGATTTCAGGC	50.5	50	CCCCATATCGTCTGTTGAA	51.5	47.4	614
	5	GTTTTGGTGTAGCTGTGAGT	57.2	50	TCGTCTGTTGAAGGCGTGC	57.8	57.9	506
	6	TGATTCGGTTTTGGTGTGGC	56.2	47.6	CAGAGTACCAATGCGGATGA	55.9	45.5	491
	9	AGGGAGCATGTTGCTAC	54.3	55.6	GATGCTTGACCGAGTCCC	55.3	61.1	262
	10	TTCTGAGGGAGCAAGTTCGC	57.1	55	TTGAAGGCGTGCACGGAAC	58.4	57.9	462
	11	CACAAGTCTCATTGCCCA	52.4	50	GTATTGTCTGTTGAAGGCG	51.2	47.4	330

Polymorphism detection in many crops were already been reported while no studies have been performed in WRKY genes in *Arecaceae*. In an earlier report in ginger EST sequences, SNPs are found with a frequency of 0.84 SNPs/100bp (Riju *et al.*, 2009) and 1.36 per 100 bp in oil palm (Riju *et al.*, 2007). SNPs (SNP=1.72 and indel=1.1) in the present study is also comes in line with the similar range of frequency. Studies had suggested that the majority of the predicted SNPs and indels represent true genetic variation in ginger. So in this study of SNP in WRKY palms, the frequency of occurrence of SNPs was found to be increased shows that the high genetic variation among the WRKY loci will be due to the presence of high polymorphism. There was a similar study of mining of SNPs in maize by Batley *et al.*, (2003), in which they have identified over 14,832 candidate SNPs in maize EST sequence, they demonstrates that candidate SNPs with high redundancy and co segregation confidence scores are likely to represent true SNPs. Also the transition to transversion ratio and indel size frequencies corresponds to those observed by direct sequencing methods of SNP discovery and suggested that the majority of predicted SNPs and indels identified using this approach represent true genetic variation in maize.

The predicted SNP data could be made valuable by the direct sequencing method in order to validate the electronic prediction. This fact was found successful in Maize EST sequence data by Batley *et al.*, 2003. In an already circulated report on an important *arecaceae* member, the *cocos nucifera L.*, WRKY SNPs are used as molecular markers to understand the diversity (Mauro-Herrera *et al.*, 2006, 2007). This study included the WRKY families, WRKY-01 WRKY-02, WRKY-03, WRKY-05, WRKY-06, WRKY-10, WRKY-13, WRKY-16, WRKY-19, and WRKY-21, but does not include other WRKY families. Recently Meerow *et al.*, (2009) reported gene sequences of seven WRKY families which we had used for mining. The preliminary report of primers amplifying WRKY loci directly from genomic DNA imparts a way for developing important

and beneficial genetic markers from members of a WRKY gene family (Borrone *et al.*, 2004).

Here the SNPs were seen at a frequency of 2.84/100bp in the WRKY sequences of palms. Primers were designed flanking to SNP/ indel sites with potential as markers in palms. PCR products from these amplicons could be resequenced to validate the detected SNPs and evaluated for use as molecular markers. We could obtain two novel electronic WRKY-SNP markers (WRKY 7 and WRKY 12) which are not reported before in palms.

### Conclusion

Potential SNP sites from the study could also prove useful to detect polymorphism in palm WRKY germplasm and also linkage mapping. The present data confirm that the frequency of SNP occurrence in palm is sufficient to make them appropriate markers for any kind of genetic studies. The two novel SNP markers and their corresponding designed primers provided by this study can lead to many mapping and polymorphic analysis. Some relevant primer pairs flanking to the desired WRKY polymorphic sites can also be used by the scientific community.

### Acknowledgment

The work was supported by a grant from Department of Information Technology (DIT), Government of India. We are thankful to Dr. George V. Thomas, Director, CPCRI, Kasaragod for the encouragement, guidance and facilities.

### References

- Batley, J., Barker, G., Helen, S., Keith, J. E., and Edwards, D. 2003. Mining for Single Nucleotide Polymorphisms and Insertions/Deletions in Maize Expressed Sequence Tag Data. *Plant Physiology* **132** (1): 84–91.
- Borrone, J. W., Kuhn, D. N. and Schnell, R, J. 2004. Isolation, characterization, and development of WRKY genes as useful genetic markers in *Theobroma cacao*. *Theor Appl Genet* **109** (3): 495–507.

- Buetow, K. H., Edmonson, M. N., and Cassidy, A. B. 1999. Reliable identification of large numbers of candidate SNPs from public EST data. *Nat Genet* **21** (3): 323–325
- Cheng, T. C., Xia, O. Y., Qian, J. F., Liu, C., Lin, Y., Zha, X. F. and Xiang, Z. H. (2004) Mining single nucleotide polymorphisms from EST data of silkworm, *Bombyx mori*, inbred strain Dazao. *Insect Biochem. Mol. Biol.* **34** (6): 523-530.
- Douglas, E. S., Pamela, S. S. and Doyle, J. J. 1998. Book- *Molecular systematics of plants 2- DNA sequencing*, Kluwer Academic publishers. United States of America. Volume 2: 556
- Eulgem, T., Rushton, P. J., Robatzek, S. and Somssich, I. E. 2000. The WRKY superfamily of plant transcription factors. *Trends in plant Science* **5** (5): 199-205.
- Garg, K., Green, P. and Nickerson, D.A. 1999. Identification of candidate coding region single nucleotide polymorphisms in 165 human gene using assembled expressed sequence tags. 1999. *Genome Res* **9** (11): 1087-92.
- Gu, Z., Hillier, L. and Kwok, P. Y. 1998. Single nucleotide polymorphism hunting in cyberspace. *Hum Mutat* **12** (4): 221–225.
- Gupta, P. K., Roy, J. K. and Prasad, M. 2001. Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Current Science* **80** (4): 524-535.
- Huang, X., and Madan, A. 1999. CAP3: A DNA Sequence Assembly Program. *Genome Research* **9**: 868-877.
- Kim, H. J., Lee, H. R., Han, J. H., Yeom, S. I., Harn, C. H., and Kim, B. D. 2008. Marker production by PCR amplification with primer pairs from conserved sequences of WRKY genes in chili pepper. *Molecules and Cells* **25** (2): 196-204.
- Lee, S. H., Park, E. W., Cho, Y. M., Ji-Woong, L., Kim, H. Y., Jun-Heon, L., Sung-Jong, O., Il-Cheong, C. and Du-Hak, Y. 2005. Confirming Single Nucleotide Polymorphisms from Expressed Sequence Tag Datasets Derived from Three Cattle cDNA Libraries. *Biochem and Mol Bio* **39** (2): 183-188.
- Lopez, C. E., Quesadao, L.M., Bohorquez, A., Duque, M. C., Vargas, J., Joe, T. and Verdier, V. 2007. Mapping EST-derived SSRs and ESTs involved in resistance to bacterial blight in *Manihot esculenta*. *Genome* **50** (12): 1078–1088.
- Marth, G. T., Korf, I., Yandell, M. D., Yeh, R. T., Gu, Z., Zakeri, H., Stitzel, N. O., Hillier, L., Kwok, P. Y. and Gish, W.R. 1999. A general approach to single-nucleotide polymorphism discovery. *Nat Genet* **23** (4): 452–456.
- Mauro, H. M., Meerow, A.W., Borrone, J. W., Kuhn, D. N. and Schnell, R. J. 2006. Ten informative markers developed from WRKY sequences in coconut (*Cocos nucifera*). *Molecular Ecology Notes* **6** (3): 904–906.
- Mauro, H. M., Meerow, A.W., Borrone, J. W., Kuhn, D. N., Schnell, R. J. 2007. Usefulness of WRKY gene-derived markers for assessing genetic population structure: An example with florida coconut cultivars. *Scientia Horticulturae* **115** (1): 19-26.
- Meerow, A. W., Noblick, L., Borrone, J. W., Couvreur, .L.P., Mauro-Herrera, M., Hahn, W. J., Kuhn, D. N., Nakamura, K., Oleas, N. H and Schnell, R. J. 2009. Phylogenetic Analysis of Seven WRKY Genes across the Palm Subtribe Attaleinae (Arecaceae) Identifies Syagrus as Sister Group of the Coconut. *PLoS ONE* **4** (10) e: 7353.
- Nikiforov, T. T., Rendle, R. B., Goelat, P, Rogers, Y. H., Kotewicz, M. L., Anderson, S., Trainor, G. L. and Knapp, M. R. 1994. Genetic bit analysis: A solid phase method for typing single nucleotide

- polymorphisms. *Nucleic Acids Research* **22** (20): 4167–4175.
- Pandey, S. P. and Somssich, I. E. 2009. The Role of WRKY Transcription factors in Plant Immunity. *Plant Physiology* **150** (4): 1648-1655.
- Picoult-Newberg, L., Ideker, T. E., Pohl, M. G., Taylor, S. L., Donaldson, M. A., Nickerson, D. A. and Boyce-Jacino, M. 1999. Mining SNPs From EST Databases. *Genome Research* **9**: 167-174.
- Riju, A., Chandrasekar, A., Sithara, K., Anoop, S., and Eapen, S. J. 2009. Identification of single nucleotide polymorphism in ginger using expressed sequence tags. *Bioinformation* **4**(3): 119-122.
- Riju, A., Rajesh, M. K., Fasila, S., Chandrasekar, A., Apshara, E., and Arunachalam, V. 2009. Mining of expressed sequence tag libraries of cacao for microsatellite markers using five computational tools. *Journal of genetics* **88** (2): 217-225.
- Taillon-Miller, P., Gu, Z.J., Li, Q., Hillier, L. and Kwok, P. Y. 1998. Overlapping genomic sequences: a treasure trove of single-nucleotide polymorphisms. *Genome Res* **8** (7): 748–754.
- Thiel, T., Michalek, W., Varsheny, R. K. and Graner, A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet* **106** (3), 411–422.
- Wu, K. L., Guo, Z. J., Wang, H. H. and Jing, Li. 2005. The WRKY family of transcription factors in Rice and *Arabidopsis* and their origins. *DNA Research* **12** (1): 9-26.
- Xie, Z., Zhong-Lin, Z., Xiaolu, Z., Huang, J., Ruas, P., Daniel, T., and Qingxi, J. S. 2004. Annotations and Functional Analyses of the Rice WRKY Gene Superfamily Reveal Positive and Negative Regulators of Abscisic Acid Signaling in Aleurone Cells. *Plant Physiology* **137** (1): 176-189.