Indian Journal of Agricultural Sciences **89** (1): 111–6, January 2019/Article
https://doi.org/10.56093/ijas.v89i1.86188

In-silico analysis of *WRKY* Transcription Factors gene family in healthy and malformed stages of mango (*Mangifera indica*)

ASHOK YADAV[1], K USHA[2] and PAWAN KUMAR JAYASWAL[3]

ICAR-Indian Agricultural Research Institute, New Delhi 110 012

Received: 19 July 2018; Accepted: 20 August 2018

ABSTRACT

WRKY proteins play crucial roles in plant defense regulatory networks, development process and physiological programs including responses to several biotic and abiotic stresses. Evalutionary analysis revealed, *WRKY* genes were categorized into the four major groups. In developed phylogenetic tree, group-D contain highest number (15) of *WRKY* genes followed by group-B (10), group-A (7), and group-C (6). Several number of CRE's were identified from mango transcriptome belonging to different categories like light responsiveness, hormone responsive, biotic stress responsive, biotic stress responsive, binding, plant development, transcription and circadian control. Among the 10 stable genes observed in transcriptome, nine genes had negative Z-score indicating that these structures identified for the proteins are reliable. Motif analysis indicated that the per cent occurrence of all the five motifs were higher in *WRKY* genes of malformed tissues compared to *WRKY* genes of healthy tissues. The uniquely identified CRE's (Healthy stages: AC-II, GCC box, OBP; Malformed stages: Aux-RR-core, AC-I, 3-AF1 binding site, CAT-box, MNF1 and rbcS-CMA7a.), defense and stress responsiveness (TC-rich repeats) and fungal elicitor (Box-W1) related *cis*-regulatory elements will provide insight to solve the problem of mango malformation. The identified information regarding the *WRKY* Transcription Factor from mango transcriptome will serve as a valuable information for mango breeding against malformation.

Key words: Cis-regulatory element, Defense, *In-silico*, Motif, Orthologs, Phylogenetic

Mango (*Mangifera indica* L.) occupies paramount place among the fruit crops grown in India. Mango malformation is a serious constraint to mango production in India and other mango growing countries of the world (Kumar *et al.* 2011). In India, it causes approximately 50-80 % fruit yield loss every year. During the development process of bud to panicle, plant undergoes highly different biochemical and physiological changes in gene expression profiles. These differences provide an ingenious system to discover molecular mechanisms and genes for mango malformation. The *WRKY* transcription factor plays a prominent role in plant biological processes, i.e. plant growth and development, adaptation to adverse climatic conditions, resistance to biotic and abiotic stresses and defense signaling (Eulgem *et al.* 2000). They may acts as a positive or negative regulators in plant for several biotic abiotic and stresses. Since hormone signaling plays a major role in different kind of stresses and it has been reported that in crosstalk of hormones, *i.e.* gibberellins, abscisic acid, jasmonic acid, salicylic acid and ethylene. *WRKY* TF's also plays a crucial role in abiotic stresses (Water, salt, low and high temperatures, drought and UV radiation) and biotic stresses (nematode, response to herbivory, wounding). Besides these they are also involved in plant development process such as seed development, dormancy and germination, regulation of plant growth, metabolic pathways and plant senescence (Rushton *et al.* 2010).

TF's (Transcription factors) are sequence-specific DNA binding sites and are able to modify the transcription rate of downstream target genes and also play vital role in regulating gene expression (Martinez 2002). Among the transcription factors, *WRKY* transcription factors fit to a large family of transcription factor which are primarily been located in plants and also characterized in several and diverse plants species. The *WRKY* genes mainly consists of domain which is around 60 amino acid residues (Eulgem *et al.* 2000). The *WRKY* domain at N-terminal mainly consists of a highly conserved amino sequence, i.e. WRKYGQK whereas at C-terminal end consists of chelating zinc finger motif. The main sequence of a *WRKY* motif is WRKYGQK, with its some others variants which are WRKYGEK, WKKYGQK, WRKYGRK, and WRKYGKK. After the first report of WRKY genes in 1994 (sweet potato), till now several genes have been found in varied range of plant species.

[1]Scientist (e mail: ashokiari1@gmail.com), ICAR-Central Institute for Subtropical Horticulture, Regional Research Station, Malda, West Bengal 732 102. [2]Division of Fruits and Horticultural Technology, ICAR-IARI, New Delhi. [3]National Research Centre on Plant Biotechnology, New Delhi 110 012.

In the preset era of omics sciences, several research projects in plant crops are running to find out cues regarding the biotic and abiotic stress. Many projects has find out the gene of interest useful in plant breeding. In all living organisms, the regulation of gene expression is a dynamic mechanism for finding out the solution several problems. Due to very meagre information of *WRKY* genes in mango, we mainly conduct this study because of involvement of *WRKY* genes in biotic and abiotic stresses. Mango malformation is complex disease in mango. So we carried out the transcriptomic analysis of healthy and malformed stages to find out correlation of *WRKY* genes with mango malformation. Therefore we preformed the *in-silico* analysis of *WRKY* genes for multiple sequence alignment, gene phylogeny, conserved motif prediction, *cis*-regulatory element prediction and physico-chemical properties of the *WRKY* genes.

MATERIALS AND METHODS

To generate transcriptome, we isolated the RNA from different flowering stages of Amrapali variety of mango which includes, i.e. three malformed and two healthy stages. The five samples used for the study are as follows: single swollen malformed bud stage 1 (MB 1), multiple malformed bud stage 2 (MB 2), multiple malformed bud stage 3 (MB-3), healthy bud stage 1(HB-1) and healthy bud stage 2 (HB-2). The next generation sequencing for MB-1, MB-2, MB-3, HB-1 and HB-2 stages were performed using 2×150PE chemistry on the Illumina NextSeq platform and approximately 5-6 GB of data was generated per sample. The predicted CDS were subjected to similarity search against NCBI's non-redundant (NR) database using the BLASTx algorithm. After annotation of different *WRKY* genes were observed and their sequences were extracted for further *in-silico* analysis.

Multiple Sequence Alignment of 38 *WRKY* genes was done by Unipro UGENE software and for similarity index threshold value was kept 70% (Okonechnikov *et al*. 2012). We developed a Bayesian phylogenetic tree of 38 different *WRKY* genes identified in healthy and malformed tissue of *Mangifera indica* with Mrbayes v 3.2.2 software (Ronquist and Huelsenbeck 2003), with a much more robust node support. Total 1000 number of tree was generated and summarize with sumtburnin=250 tree (25% of total developed tree). The developed consensus Bayesian phylogenetic tree was well resolved with posterior probability (PP) varied from 0.53 to 1. The developed consensus tree is visualized with Fig tree (http://tree.bio.ed.ac.uk/software/figtree/; date of access: Jan 19, 2017) software.

The nucleotide motifs of *WRKY* genes were obtained using multiple expectation maximization for motif elicitation suite 4.10.2 (http://memesuite.org) software and the motif alignment and search tool (MAST). To assess the functional motifs of annotated amino acid sequences, E-value was set 0.01 and motif length from 12 to 60. The 1 kb upstream sequences of *WRKY* genes were used to find cis- regulatory elements in the PlantCARE database (Lescot *et al*. 2002).

ProtParamExPASy(Gasteiger *et al*. 2005) tool was used for calculating various physiochemical parameters **(**molecular weight, theoretical pI, GRAVY and instability index) of *WRKY* genes. Protein sequences of *WRKY* genes were used as an input source. Molecular weight of protein is calculated by adding the average isotopic masses of amino acids (provided protein) and the average isotopic mass of one water molecule. Protein pI is calculated using pKa values of amino acids. The GRAVY value for a protein was obtained by adding the hydropathy values of each amino acid residues and dividing by the number of residues in the length of the sequence. A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable. The protein sequences of *WRKY* genes were used to build 3D models by the phyre2 server. Further validation of protein structures was carried out from X-ray analysis, NMR spectroscopy and Z-score estimation was done by ProSA-web tool (Wiederstein and Sippl 2007).

RESULTS AND DISCUSSION

Protein sequences of *WRKY* genes were used to find out the homologous sequences which will be helpful in developing marker and *Fusarium* resistance genes for mango malformation and other horticultural disease. The multiple sequence alignment of polymorphic sites in *WRKY* genes showed the conserved region. The phylogenetic tree was divided into four major groups, namely A, B, C and D shown with different color. The group-D contain highest

Table 1 Details of five best motifs present in WRKY genes

Motif	Motif width	Sites	E-value	Best possible match
Motif-1	60	38	7.7e-802	TGGMGNAAATATGGACARAAASTWGTVAAAGGMAATCCHAATCCAAGGAGCTAYTACAAG
Motif-2	60	34	3.9e-599	GRGCATCHCAYGATCYRARRDCDGTKATCACMACHTATGARGGRAARCACAACCATGATG
Motif-3	60	32	7.2e-569	GAGCCWAGAGTTGTDGTTCAAACAACAAGTGAWRTTGAYATTCTTGATGATGGATAYMGC
Motif-4	60	27	1.1e-526	TTGGAGAAAATATGGRCAAAAACAAGTSAAAGGAAGTGARWATCCWMGRAGYTAYTACAA
Motif-5	60	32	4.6e-415	WGRWGATGATGNTGAHGAARATGAAYCBGAGTCMAARAGAWGGAAAANDGAVRNTGAWRH

number (15) of *WRKY* genes followed by group-B (10), group-A (7), and group-C (6). Among the four groups identified in phylogenetic tree, *WRKY-22* genes (group-C) were recently evolved genes and these genes were diverged from *WRKY-33* genes (group-D). The *WRKY-2* and *WRKY-33* genes were parallely evolved from *WRKY-1* genes, whereas the *WRKY-1* genes were the most ancestral genes. Multiple sequence alignment (MSA) and phylogenetic tree construction are becoming powerful tool in plant science for biological function analysis and performing the task of next-generation sequencing (Ortuno *et al.* 2013). Similarly phylogenetic analysis of *WRKY* genes has been performed by Huang *et al.* (2015) in white pear, Li *et al.* (2016) in carrot and Zhang *et al.* (2017) in potato.

In-silico motif analysis of 38 *WRKY* genes for motif prediction resulted five significant motifs with minimum and maximum width of 12 and 60 respectively were mined and are designated as motifs 1, motif 2, motif 3, motif 4 and motif 5 (Table 1; Fig 2). Among the all *WRKY* genes studied, motif-1 (Red color) was present in all 38 genes followed by motif-2 (Blue color) in 34 genes, motif-3 (Light green color) in 32 genes, motif-5 (yellow color) in 32 genes and motif-4 was present in 27 genes. The per cent occurrence of all the five motif (71.42, 90.47, 85.71, 71.42 and 85.71) were higher in *WRKY* genes of malformed tissue compared to *WRKY* genes of healthy tissue (70.58, 88.24, 82.35, 70.58 and 82.35). Zhang *et al.* (2009) correlated the statistical significance of predicted motif with biological significance which gave a valuable results. Several study on motif prediction have resulted diverse application of motif for crop improvement such as gene expression analysis study (Jensen *et al*. 2005, Huber and Bulyk 2006), discovery of sub-families in large protein families (Leonardi and Galves 2005), family classification (Blekas *et al*. 2005, Eser *et*

Table 2 Grouping of identified *cis*-regulatory elements in functional categories

Functional categories	Type of *cis*-regulatory element
Light responsiveness CRE's	ACE, ATC-motif, ATCT-motif, AE-box, Box I, Box II, Box 4,CHS-CMA2a, CHS-Unit 1 m1, CATT-motif, GATA-motif, as-2-boxGA-motif, G-box, GATT-motif,GAG-motif, GTGGC-motif, GT1-motif, I-box, LAMP-element, MNF1, MRE, Pc-CMA2c, rbcS-CMA7a, Sp1, TCCC-motif, TGG-motif, TCT-motif, 3-AF1 binding site
Hormone responsive CRE's	Abscisic acid responsive: ABRE
	Auxin-responsive: TGA-element, AuxRR-core
	Ethylene-responsive: ERE
	Gibberellin-responsive: P-box, GARE-motif, TATC-box
	Salicylic acid responsive: TCA-element
Biotic stress responsive CRE's	Defense and stress responsiveness: TC-rich repeats
	Fungal elicitor: Box-W1
	Elicitor-responsive: EIRE
	Maximal elicitor-mediated activation: AT-rich sequence
	MeJa-responsiveness: CGTCA-motif, TGACG-motif
Abiotic stress responsive CRE's	Heat stress: HSE
	Anaerobic induction: ARE
	Drought-inducibility: MBS
	Anoxic specific inducibility: GC-motif
	Low-temperature responsiveness: LTR
Plant development related CRE's	Zein metabolism regulation: O_2-site
	Endosperm expression: Skn-1_motif
	Endosperm expression: GCN4_motif
	Meristem expression: CAT-box
CRE's involved in binding	Protein binding: CCAAT-box, Box III
	Mybhv1 binding site: CCAAT-box
	DNA binding protein : OBP1 site
	AT-rich DNA binding protein (ATBP-1): AT-rich element
Conferring high transcription	5UTR PY-rich stretch
Involved in circadian control	Circadian
CRE's in promoter and enhancer regions	CAAT-box
Mediating transactivation by MYB transcription factors during lignin biosynthesis	AC-I
Core promoter element around -30 of transcription start	TATA-box
Unknown Function	AC-II, AAGAA-motif, Box S, Box E, G-box, GCC box, TCCACCT-motif, TATCCAT/C-motif, Un-named_1, 2, 3, 4,5, 8, 9, 10, 11,12, 13,14 and 17, W-box

al. 2013), new signalling pathways (Ma *et al*. 2013) and can be used for developing resistance genes and makers (Broin *et al.* 2015), and discovery of homology relations (Stewart 2016).

Analysis of 38 *WRKY* genes resulted in 82 types of *cis*-regulatory elements. The functions of different *cis*-regulatory elements are described in Table 2. The unique cis-regulatory element in malformed stages were AC-II, GCC box and OBP-1 site, whereas in healthy stages were Aux-RR-core, AC-I, 3-AF1 binding site, CAT-box, MNF1 and rbcS-CMA7a (Table 3). According to differences in function, the identified *cis*-regulatory elements were classified into 11 categories along with unknown function category. The highest percentage of CRE's were observed in light responsiveness category (38 %) followed by hormone responsive (9 %), biotic stress responsive (7%), abiotic stress responsive (6%), binding (6%), plant development related CRE's (5%), conferring high transcription (1%), involved in circadian control (1%), CRE's in promoter and enhancer regions (1%), mediating transactivation by MYB TF (1%), core promoter element (1%) whereas 28% CRE'S were having unknown function. The uniquely identified CRE's, defense and stress responsiveness (TC-rich repeats) and fungal elicitor (Box-W1) related *cis*-regulatory element will be helpful in providing insight for solving the mango malformation problem. Similarly, Kaur *et al.* (2017**)** identified CRE's from pathogenesis-related proteins of *Arabidopsis thaliana* and *Oryza sativa.*

Protein sequence analysis in 38 *WRKY* genes retrieved from mango transcriptome of flowering stages indicated that in malformed stages maximum molecular weight was observed in *MB2_WRKK2_2* (179741.05 kDa) while minimum in *MB1_WRKK22*_2, i.e. 48369.93 kDa, whereas in healthy stages it was maximum (157744.98 kDa) *HB1_WRKK2_1* and minimum (51584.89 kDa) in *HB1_WRKK22_2* (Table 4).The observed *pI* of *WRKY* genes in malformed stages resulted maximum (5.15) in *MB1_WRKK22_1* and minimum in *MB2_WRKK2_1* (4.92), whereas in healthy stages it was maximum in *HB1_WRKK22_2* (5.16) and minimum in *HB2_WRKK33_1* (4.89). Our results indicated that among 21 *WRKY* genes of malformed stage five were stable (23.80%) while in healthy stages five genes were stable out of 17 *WRKY* genes (29.41%). At single bud stage (MB-1 vs HB-1) and multiple bud stages (MB-2 vs HB-2) the number of stable *WRKY* genes were more in healthy stages compared to malformed stages. The grand average of hydropathicity (GRAVY) ranged from 0.749 (*MB3_WRKK2_3*) to 0.999 (*MB3_WRKK22_1*) in malformed stages while in healthy stages it varied from 0.770 (*HB2_WRKK2_2*) to 0.956 (*HB2_WRKK22_1*) (Table 4). The information generated on different protein parameters of *WRKY* genes can be used in identification of homologs related to *WRKY* genes in other plant genomes, preservation of the genetic code, and also to measure the hydrophobicity of a specific peptide/protein related to malformation resistance. Similarly physico-chemical properties for protein sequences were observed by *in-silico* analysis in different plant crops (Crasto 2010, Sahay and Shakya 2010, Mallikarjuna *et al.* 2016).

On the basis of instability index, out of 38 *WRKY* genes 10 stables genes were selected for secondary and tertiary structures analysis. The secondary structure mainly consists of alpha helix, beta strand, disordered and TM helix. The secondary structure analysis suggested that alpha helix percentage was highest (15%) in *HB1_WRKK2_2* genes of healthy stages while minimum in *HB2_WRKK1_1*gene of healthy bud stage (7%). The beta strand percent was more in *HB2_WRKK1_1* (10%) and minimum in *HB1_WRKK2_2* (7%). The disordered percent in *WRKY* genes of healthy stages varied from 59 to 82 %, whereas in *WRKY* genes of malformed stages varied from 59 to 67 %. Among the all 10 stables genes of healthy and malformed stages only one gene, i.e. *MB3_WRKK2_3* had three transmembrane helix (Table 5). This information of secondary structure (alpha helix, beta strand and disordered percentage) of protein will be helpful in understanding both the mechanisms of folding and the biological activity of proteins (Sivan *et al.* 2007).The Z-score observed among the 10 stable genes was negative in all genes except in *HB1_WRKK1_2* (3.11). The negative value of Z-score obtained for different *WRKY* genes indicated that these structures are reliable. Similarly, Prajapat *et al*. (2007) in AC1 proteins of begomo virus strains and Mishra *et al.* (2015) in chitinase gene family of wheat obtained negative value of Z score.

The information generated through phylogenetic trees, *cis*-acting elements and motif prediction obtained from present study will provide better insights of the

Table 3 List of unique cis-regulatory element observed in malformed and healthy stages

Cis-regulatory element	Sequences	Function
Malformed stage		
AC-II	TCAACCAACTCC	Unknown
GCC box	AGCCGCC	Unknown
OBP-1 site	TACACTTTTGG	*cis*-acting regulatory element
Healthy stage		
3-AF1 binding site	TAAGAGAGGAA	Light responsive element
Aux-RR-core	GGTCCAT	Involved in auxin responsiveness
CAT-box	GCCACT	Meristem expression
MNF1	GTGCCC(A/T)(A/T)	Light responsive element
rbcS-CMA7a	GTCGATAAGG	Light responsive element
AC-I	GCTTACCTACCA	Mediating transactivation by MYB transcription factors during lignin biosynthesis

Table 4 Physico-chemical properties of the identified *WRKY* genes in healthy and malformed tissue

Stage	Genes	Number of amino acids	Molecular weight	pI	Instability index	Aliphatic index	GRAVY
MB-1	MB1_WRKK1_1	1509	122717.27	5.03	34.52	34.19	0.805
	MB1_WRKK1_2	1509	122717.27	5.03	34.52	34.19	0.805
	MB1_WRKK2_1	1047	85586.62	5.07	47.14	28.18	0.779
	MB1_WRKK22_1	606	51456.18	5.15	53.28	29.04	0.857
	MB1_WRKK22_2	585	48369.93	5.14	55.80	28.38	0.923
	MB1_WRKK33_1	1758	146818.62	4.93	49.44	31.51	0.944
	MB1_WRKK33_2	1737	145123.70	4.93	48.04	31.15	0.939
	MB1_WRKK33_3	1680	140418.43	4.96	51.63	30.89	0.878
	MB1_WRKK33_4	1737	145256.82	4.95	52.00	30.69	0.871
MB-2	MB2_WRKK1_1	1497	121446.05	5.03	33.27	34.80	0.830
	MB2_WRKK2_1	2181	178820.93	4.92	42.85	30.22	0.808
	MB2_WRKK2_2	2190	179741.05	4.92	43.12	30.09	0.809
	MB2_WRKK33_1	1776	148514.55	4.93	49.30	31.36	0.941
	MB2_WRKK33_2	1737	145228.82	4.95	52.30	30.74	0.874
MB-3	MB3_WRKK1_1	1509	122673.31	5.03	33.18	33.73	0.806
	MB3_WRKK2_1	1047	85552.52	5.07	47.17	47.17	0.775
	MB3_WRKK2_2	1038	84586.32	5.07	46.36	28.52	0.769
	MB3_WRKK2_3	1677	138210.74	4.99	37.00	29.93	0.749
	MB3_WRKK22_1	696	59259.96	5.09	63.27	29.45	0.999
	MB3_WRKK33_1	1725	144210.58	4.93	45.08	30.43	0.925
	MB3_WRKK33_2	1737	145228.82	4.95	51.86	30.74	0.874
HB-1	HB1_WRKK1_1	1509	122717.27	5.03	34.52	34.19	0.805
	HB1_WRKK1_2	1497	121398.06	5.03	34.22	35.20	0.839
	HB1_WRKK2_1	1929	157744.98	4.95	41.51	30.59	0.795
	HB1_WRKK2_2	1200	99613.19	4.97	37.49	29.92	0.772
	HB1_WRKK22_1	687	58643.75	5.11	59.44	28.68	0.912
	HB1_WRKK22_2	615	51584.89	5.16	53.56	28.46	0.792
	HB1_WRKK33_1	1746	145917.49	4.93	46.32	30.87	0.929
	HB1_WRKK33_2	1734	144906.40	4.93	47.43	31.03	0.934
	HB1_WRKK33_3	1716	143523.66	4.96	50.40	30.77	0.859
	HB1_WRKK33_4	1662	138854.45	4.96	49.89	31.11	0.869
	HB1_WRKK33_5	1737	145212.76	4.95	52.23	30.69	0.871
HB-2	HB2_WRKK1_1	1509	122867.82	5.02	34.33	33.73	0.824
	HB2_WRKK2_1	1380	112285.44	5.01	43.55	29.49	0.801
	HB2_WRKK2_2	1077	89741.24	4.98	36.66	29.62	0.770
	HB2_WRKK22_1	903	76714.64	5.05	62.32	28.90	0.956
	HB2_WRKK22_2	1755	146455.19	4.93	48.77	31.51	0.944
	HB2_WRKK33_1	1734	145019.66	4.89	52.44	30.62	0.876

transcriptional gene regulation system. It is essential to decipher the expression of these resistance genes, *cis-regulatory* elements and markers in economically important horticultural crops to improve disease resistance. The 3D structure of ten stable *WRKY* genes can be effectively used for *in silico* docking study for development of potential ligand molecules against *Fusarium* infection. Potential ligand molecules against *Fusarium mangiferae* infection can be developed by *in silico* docking from 3D structure of 10 stable *WRKY*genes. The present study on *in-silico* analysis of *WRKY* genes in healthy and malformed tissue can be used to improve resistance against mango malformation through genome editing and *gene* silencing strategies.

ACKNOWLEDGEMENT

The financial assistance provided by Director, ICAR-IARI, New Delhi, India for conducting the studies is gratefully acknowledged.

Table 5 Secondary structure analysis and Z-score prediction of WRKY genes

Genes	Disordered	Alpha helix	Beta strand	TM helix	Z-Score
MB1_WRKK1_1	60	11	8		-3.06
MB1_WRKK1_2	60	11	8		-3.56
MB2_WRKK1_1	59	11	9		-2.41
MB3_WRKK1_1	63	10	9		-3.87
MB3_WRKK2_3	67	12	9	3	-3.05
HB1_WRKK1_1	60	11	8		-3.66
HB1_WRKK1_2	60	12	9		3.11
HB1_WRKK2_2	82	15	7		-2.65
HB2_WRKK1_1	59	7	10		-3.01
HB2_WRKK2_2	78	15	8		-3.44

REFERENCES

Blekas K, Fotiadis D I and Likas A. 2005. Motif-based protein sequence classification using neural networks. *Journal of Computational Biology* **12**: 64–82.

Bockus W W, Bowden R L, Hunger R M, Morrill W L, Murray T D and Smiley R W. 2007. *Compendium of Wheat Diseases and Insects*, 3rd Edition. APS Press, St. Paul, MN.

Broin P Ó, Smith T J and Golden A A. 2015. Alignment-free clustering of transcription factor binding motifs using a genetic-k-medoids approach. *BMC Bioinformatics* **16**:22.

Crasto C J. 2010. Hydrophobicity profiles in G protein-coupled receptor transmembrane helical domains. *Journal of Receptor, Ligand and Channel Research:* 123.

Eser E M, Arslan B R and Sezerman U O. 2013. Classification of cohesin family using class specific motifs. (*In*) *Health Informatics and Bioinformatics (HIBIT)*, 8th International Symposium on 25 Sep 2013 IEEE.

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): 199–206.

Gasteiger E, Hoogland C, Gattiker A Duvaud S, Wilkins M R, Appel R D and Bairoch A. 2005. Protein Identification and analysis tools on the ExPASy Server. (*In*) *The Proteomics Protocols Handbook,* pp 571–60. John M. Walker (Ed). Humana Press.

Huang X, Li K, Xu X, Yao Z, Jin C. and Zhang S. 2015. Genome-wide analysis of WRKY transcription factors in white pear (*Pyrusbret schneideri*) reveals evolution and patterns under drought stress. *BMC genomics* **16**(1): 1104.

Huber B R and Bulyk M L. 2006. Meta-analysis discovery of tissue-specific DNA sequence motifs from mammalian gene expression data. *BMC Bioinformatics* **7:** 229.

Jensen S T, Shen L, and Liu J S. 2005. Combining phylogenetic motif discovery and motif clustering to predict co-regulated genes. *Bioinformatics* **21**: 3832–9.

Kaur A, Pati P K, Pati A M and Nagpal A K. 2017. *In-silico* analysis of cis-acting regulatory elements of pathogenesis-related proteins of *Arabidopsis thaliana* and *Oryza sativa*. *PloS one* **12** (9): e0184523.

Kumar P, Misra A K and Modi D R. 2011.Current status of mango malformation in India. *Asian Journal of Plant Sciences* **10**: 1.

Leonardi F and Galves, A. 2005. Sequence motif identification and protein family classification using probabilistic trees. (*In*) *Brazilian Symposium on Bioinformatics*. Springer, Berlin, Heidelberg, pp 190–3.

Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P and Rombauts S. 2002. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research* **30** (1): 325–7.

Li M Y, Xu Z S, Tian C, Huang Y, Wang F and Xiong A S. 2016. Genomic identification of WRKY transcription factors in carrot (*Daucus carota*) and analysis of evolution and homologous groups for plants. *Scientific Reports* **6**: srep23101.

Ma S, Shah S, Bohnert H J, Snyder M and Dinesh-Kumar S P. 2013. Incorporating motif analysis into gene co-expression networks reveals novel modular expression pattern and new signaling pathways. *PLoS Genet* **9**: e1003840

Mallikarjuna M G, Nepolean T, Mittal S, Hossain F, Bhat J S, Manjaiah K M, Marla S, Mithra A C, Agrawal P K, Rao A R and Gupta H S. 2016. In-silico characterisation and comparative mapping of yellow stripe like transporters in five grass species. *Indian Journal of Agricultural Sciences* **86**: 621–7.

Martinez E. 2002. Multi-protein complexes in eukaryotic gene transcription. *Plant Molecular Biology* **50** (6): 925–47.

Mishra A K, Pandey B, Tyagi C, Chakraborty O, Kumar A and Jain A K. 2015. Structural and functional analysis of chitinase gene family in wheat (*Triticum aestivum*). *Indian Journal of Biochemistry and Biophysics* **52**: 169–78.

Ortuno F M, Valenzuela O, Pomares H, Rojas F, Florido J P, Urquiza J M and Rojas I. 2013. Predicting the accuracy of multiple sequence alignment algorithms by using computational intelligent techniques. *Nucleic Acids Research* **41**: e26.

Okonechnikov K, Golosova O, Fursov M and UGENE team. 2012. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**(8): 1166–7.

Ronquist F and Huelsenbeck J P. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–4.

Prajapat R, Marwal A and Gaur R K. 2014. Recognition of errors in the refinement and validation of three-dimensional structures of AC1 proteins of begomovirus strains by using ProSA-Web. *Journal of Viruses.*

Rushton P J, Somssich I E and Ringler P. 2010. WRKY transcription factors. *Trends in Plant Science* **15** (5): 247–58.

Sahay A and Shakya M. 2010. *In silico* analysis and homology modelling of antioxidant proteins of spinach. *J. Proteomics Bioinform* **3**: 148–54.

Sivan S, Filo O and Siegelmann H. 2007. Application of expert networks for predicting proteins secondary structure. *Biomolecular Engineering* **24**: 237–43.

Stewart Jr C N. 2016. *Plant Biotechnology and Genetics: Principles, Techniques and Applications,* p 432. John Wiley & Sons.

Timothy L B, James J, Charles E G and William S N. 2015. The MEME Suite. *Nucleic Acids Research* **43** (W1): W39–49.

Zhang C, Wang D, Yang C, Kong N, Shi Z, Zhao P, Nan Y, Nie T, Wang R, Ma H and Chen Q. 2017. Genome-wide identification of the potato *WRKY* transcription factor family. *PloS one* **12** (7): e0181573.

Zhang S, Su W and Yang J. 2009. ARCS-Motif: discovering correlated motifs from unaligned biological sequences. *Bioinformatics* **25**: 183–9.