Int.J.Curr.Microbiol.App.Sci (2018) *7*(2): 3052-3070

International Journal of Current Microbiology and Applied Sciences
***ISSN: 2319-7706* Volume 7 Number 02 (2018)**
Journal homepage: http://www.ijcmas.com

Original Research Article https://doi.org/10.20546/ijcmas.2018.702.369

Molecular Characterization and Evolutionary Analysis of Potential *Fusarium* Resistant Genes for Crop Improvement

Ashok Yadav[1*], Usha Kalidindi[1], Pawan Kumar Jayaswal[3], Shikha Mittal[2] and Nepolean Thirunavukkarasu[2]

[1]Divison of Fruits and Horticultural Technology, Indian Agricultural Research Institute, New Delhi-110012, India
[2]Division of Genetics, Indian Agricultural Research Institute, New Delhi-110012, India
[3]National Research Centre for Plant Biotechnology, Indian Agricultural Research Institute, New Delhi-110012, India

**Corresponding author*

Keywords

Fusarium, Candidate genes, *cis*-regulatory, Disease resistance protein structure

Article Info

Accepted:
26 January 2018
Available Online:
10 February 2018

A B S T R A C T

Fusarium species cause considerable crop losses worldwide and necessitate an expensive use of chemical pesticides. Resistant genes can be used to improve plant resistance through gene editing and *gene* silencing strategies. Currently, information on host defences against *Fusarium* species is limited. Through *in-silico* tools, identification of genes and understanding of disease resistance mechanism has become possible. In the present study, a comprehensive *in-silico* analysis of 65 *Fusarium* resistance genes belonging to 10 different species was carried out to understand the physiochemical properties, secondary and tertiary protein structures, protein binding sites, phylogenetic relationship, cis-regulatory elements and protein motifs. Five genes on the basis of their stability have been selected for 3D structure prediction using Phyre2 server. These structures were evaluated and validated using Verify-3D, PROCHECK and Ramachandran plot. The Z score predicted by ProSA-web tool indicated reliability of the 3D structures while above 97% of the region falls in allowed region of Ramachandran plot. The distribution pattern in five genes for binding site is high, while hydrophobicity is very low to low. The most commonly occurring motifs among *Fusarium* resistance genes were motif-3 and motif-1, whereas *cis*-regulatory elements were TC-rich repeats, Box-W1, EIRE, AT-rich sequence, CGTCA-motif and TGACG-motif. The results of *cis*-acting regulatory elements and motif analysis revealed that they are actively involved in the expression and regulation of *Fusarium* resistance genes during cellular development and under biotic stress conditions in different plant species. *In-silico* analysis revealed the homology of *Fusarium* resistance genes and will be helpful to develop molecular markers or gene specific markers to overcome the *Fusarium* related biotic stresses.

Introduction

Fusarium species are widely distributed in soil, subterranean and aerial plant parts, plant debris, and other organic substrates (Nelson *et al.*, 1994). More than 50 species of *Fusarium* have been identified, but a few cause destructive diseases like blights of

cereal crops, wilts and rots of important fruit and vegetable crops and disorders of many ornamental plants resulting in 30 to 70% yield losses(Bentley *et al.,* 2006; Bockus *et al.,* 2007; Saremi *et al.,* 2010; Chakraborty and Newton, 2011). Very little information is available regarding host defences against *Fusarium* species. *Fusarium* resistance genes are highly polymorphic and have wide recognition specificities and are commonly clustered in the genome. These genes originated through gene duplication are continuously evolving through unequal exchange and constant selection pressure by pathogen evolution (Joshi and Nayak, 2013). Conventional breeding approaches to develop plant resistance to *Fusarium* are slow due to a highly complex mechanism. Biotechnology serves as a powerful tool for manipulating disease resistance in crop plants through QTL mapping and genetic transformation with the aim of increasing crop productivity in the affected areas.

To identify homologs in resistance genes, similarity searching plays a crucial role because the sequences which are similar can be treated as homologous as they could have a same ancestor. Cis regulatory element (CRE) is region of non-coding DNA which regulates the transcription of nearby genes. *Cis*-regulatory module is a stretch of DNA, usually 100-1000 DNA base pairs in length(Davidson, 2006), where a number of transcription factors can bind and regulate expression of nearby genes and regulate their transcription rates. CREs are vital components of genetic regulatory networks, which in turn control morphogenesis, the development of anatomy and other aspects of developmental biology (Fiedler and Rehmsmeier, 2006; Ibraheem *et al.,* 2010; Hernandez-Garcia and Finer, 2014).The plant gene or resistance gene sequences through phylogenetic analysis can provide the most meaningful insights to researchers in different diverse fields, including ecology, molecular biology and physiology. Phylogenetic analysis framework resulted in evolution of various morphological and chemical traits (Ronsted *et al.,* 2012), production of secondary metabolites (Agrawal and Fishbein 2008; Agrawal *et al.,* 2009; Van den Bergh, 2016) along with various complex pathways such as N_2-fixing symbioses (Werner *et al.,* 2014), chemical defence mechanisms (Agrawal, 2011), and mustard oil production (Van den Bergh, 2016).For the evolutionary synthetic biology, construction of phylogenetic trees would serve as the foundation for better understanding the evolution of cellular pathways, macromolecular machines and other emergent properties of early life (Gaucher *et al.,* 2010). Hence, phylogenetic tree analysis would help guide organisms in forming their unique biological traits by regulating the composition of structural sequence elements.

The transcriptional regulation of genes is governed by the *cis*-elements during plant growth and development and under abiotic stress conditions (Nakashima *et al.,* 2009). *Cis*-regulatory elements are useful in crop improvement for developing stable resistance genes that are transcribed in predictable ways and can be maintained during the natural processes of reproduction of the plant. Most of the CRE's are related to "Light responsive" followed by "Hormone Responsive", "abiotic stress", "biotic stress", "binding" and "transcription process". Currently, due to availability of comprehensive modern protein sequence databases, 80% or more of metagenomic sequence have similarity in protein sequences. Similarity searching tools that provide reliable results are BLAST, PSI-BLAST, SSEARCH, FASTA and the HMMER3 programs (Johnson *et al.,* 2010).Accurate sequence alignments of resistance genes can be built after the homolog's search through multiple sequence alignments, which can serve as a base for

phenotype prediction and evolutionary analysis. Several reports on *in-silico* analysis indicate that this study is a very promising approach (Dutt, *et al.,* 2010; Mallikarjuna *et al.,* 2016; Bhati *et al.,* 2016; Azam *et al.,* 2017). In the present study, a comprehensive *in-silico* analysis of *Fusarium* resistant genes was carried out by working on multiple sequence alignment analysis, construction and visualisation of phylogenetic trees, motif prediction and physicochemical analysis of protein sequences. The information generated will be useful to develop resistant genotypes in different crops through development of molecular markers or gene specific markers.

Materials and Methods

Sequence retrieval of *Fusarium* resistance genes

The nucleotide and protein sequences of *Fusarium* resistance genes of banana (*Musa acuminata*), tobacco (*Nicotiana alata*), petunia (*Petunia x hybrid*), lilium (*Lilium regale*), onion (*Allium cepa*), soyabean (*Glycine max*), rice (*Oryza sativa*), Arabidopsis (*Arabidopsis thaliana*), tomato (*Solanum lycopersicon*) and melon (*Cucumis melo*) were collected from NCBI (www.ncbi.nlm.nih.gov) and coding sequences from European Nucleotide Archieve (http://www.ebi.ac.uk/ena/data/view /<accession>). The selected species comprised the representatives from ten main plant groups such as monocots, dicots and lower plants (Table1). Details of *in-silico* analysis carried out are shown in Figure 1.

Physico-chemical properties of *Fusarium* resistance genes

ProtParam, a protein analysis tool was used for identification of the physicochemical properties of the proteins (http://www.expasy. org/tools/protparam.html). Physiochemical properties including molecular weight, theoretical pI, grand average of hydropathicity (GRAVY) and instability index were computed using protein sequences. In ProtParam, the average isotopic masses of amino acids and the average isotopic mass of one water molecule were added to get the molecular weight of provided protein. The pI of protein is the pH at which the protein does not carry a net-charge and its value is obtained by the number of protonatable and deprotonatable side chains (Garcia-Moreno, 2009). Protein pI was calculated using pKa values of amino acids. pKa values depend on its side chain and has role in defining the pH dependent characteristics of a protein. The GRAVY value of a protein or a peptide was calculated by adding the hydropathy values of each amino acid residues and dividing by the number of residues in the sequence or length of the sequence. Positive score indicates a greater hydrophobicity. A protein whose instability index is smaller than 40 were predicted as stable and a value above 40 were predicted as unstable.

Secondary and tertiary structural analysis of protein sequences

The protein sequences of *Fusarium* resistance genes were used to build 3-D models by the phyre2 server and were evaluated using Verify-3D from PROCHECK software (Luthy, *et al.,* 1992). Ramchandran plot analysis was done by Rampage. Further validation of protein structures was carried out using ProSA web tool (Wiederstein, and Sippl, 2007; Prajapat *et al.,* 2007).

Protein binding sites and hydrophobicity analysis of protein sequence

Binding sites and hydrophobicity of five stable *Fusarium* resistance genes were predicted by phyre2 server tool and for further visualization of output PDB files was done by Discovery studio 3.5 visualizer tools (Nottensteiner, 2015).

Conserved domain search and trans-membrane prediction

The conserved domains among the five stable proteins were predicted through the CD-Search tool from NCBI's interface via RPS-BLAST (Marchler *et al.,* 2014). Trans-membrane helix was predicted through TMHMM Server (Trans-membrane Helices Hidden Markov Model (Krogh *et al.,* 2001) and visualization of trans-membrane helix was done by Protter tool (Omasits *et al.,* 2013).

Multiple sequence alignment analysis, construction and visualisation of phylogenetic trees

Multiple sequence alignment of the *Fusarium* resistance genes was carried out using CLUSTALW (http://www.genome.jp/tools/ clustalw/). The aligned sequences were subjected to the MEGA software for construction of phylogenetic tree using the NJ algorithm (Tamura *et al.,* 2013).

The Poisson correction robustness of clustering was checked by bootstrapping of 1000 replicates. Bootstrapping was used to evaluate the degree of support for a particular grouping pattern in the Phylogenetic tree. The visualization of the Phylogenetic tree was done by using Fig tree. The evolutionary distances were computed using the Composite Likelihood method (Gao and Song, 2010) and expressed as number of base substitutions per site.

Motif prediction

Protein sequences of *Fusarium* resistance genes were analysed through MEME software and the Motif Alignment and Search Tool (MAST) (http://meme.sdsc.edu/meme/website /intro. html) for motif prediction. Consensus sequences were depicted using the WebLogo version 2.8.2 (http://weblogo.berkeley.edu/).

Cis-regulatory element predictions

To find out the *cis*-regulatory elements, 1kb upstream sequence of all genes was retrieved from PhytozomeBiomart. The resistance gene sequences were searched against CRE's in the Plant CARE database, using the search for care program (Lescot *et al.,* 2002).

Results and Discussion

Physicochemical properties of *Fusarium* resistance genes

Protein physiochemical analysis indicated that among the 65 protein sequences, molecular weight of the protein ranged from a minimum of 5710.40 (I-2 clone Heamsona) to a maximum of 144827.10 (I-2 complete cds). The pIof protein value ranged from 4.17 to 11.12 (Table 1). Overall, 78.46 per cent fusarium genes were found slightly acidic in nature whereas remaining 21.54 per cent were basic in nature with a pI of ≥ 8. The instability index was minimum in CHS-7 and maximum in I-2-clone Heamsona (Table 1). Out of 65 sequences, 32 protein sequences were unstable, whereas 33 protein sequences were stable in nature. The grand average of hydropathicity (GRAVY) in the 65 *Fusarium* resistance sequences varied from -0.507 (NBS-16) to 0.898 (CHS-7).

Secondary and tertiary structural analysis of protein sequences

To analyse secondary and tertiary structures of fusarium proteins, 5 proteins (beta-glucanase, I-2-clone-2942, I-7, chit-42 and NBS-15) were selected from 65 proteins on the basis of their high stability belonging to three plants (Tomato, banana and soya bean)and one fungal species (*Trichodarma harzianum*). The secondary structure composed of alpha helix, beta stand, disordered and TM helix. The secondary structure analysis suggested that

alpha helix percentage was highest (52%) in NBS-15 gene (banana) while minimum in I-7 (20%).

The beta strand varied from 10% (I-2, & I-7) to 24% in beta-glucanase whereas disordered percentage varied from 15 to 27 % (Table S1).

The 3-D structure (Fig. 2a) predicts the arrangement of secondary structures as well as their side chains into three-dimensional space. The Ramachandran plot consisted of three regions *i.e.*, favoured region, allowed region and outlier region (Fig. 2b).

The favoured region (97.4%) was maximum in Chit-42 followed by beta-glucanase (94.2%), NBS-15 (85.2 %), I2-Clone-2942 (84.7 %) and I-7 (76.7 %) whereas allowed region was highest in I-7 (15.2 %) followed by NBS-15 (9.7 %), I-2 (8.9%), Beta-glucanase (5.8 %) and Chit-42 (2.1%) (Table S2).The Z score predicted by ProSA-web tool inI-2, I-7, NBS-15, chit-42 and B-glucanase were -7.35, -7.66, -4.68, -3.30 and -2.95 respectively (Table S3) and were obtained by X-ray crystallography (light blue) as well as NMR spectroscopy (dark blue) in five stable protein sequences of *Fusarium* resistance genes (Fig. 2c). Whereas according to Verify-3D, in beta-glucanase, 95.08% residues had 3D-1D profile >=0.2 followed by 89.60 % residues in chit-42 and 76.09 % in I-7.

Protein binding sites and hydrophobicity analysis of protein sequence

Protein binding sites and hydrophobicity level of the five genes *i.e.* beta-glucanase, I-2-clone-2942, I-7, chit-42 and NBS-15 has been represented in the Figure 3. The binding sites, hydrophobicity and hydrophilic nature have been represented by green, brown and blue colour respectively. The distribution pattern in five genes for binding site is high as compared to hydrophobicity which was very low to low.

Conserved domain search and Trans-membrane prediction

The conserved domain analysis revealed one conserved domain in each of five stable *Fusarium* resistance genes except Chit-42 and I-7 which had two domains almost similar in function. Among the genes studied for CD-search, B-glucanase, I-2-2942 andNBS-15 had Glyco-hydro-1 superfamily, P-loop NTPase superfamily and NB-ARC domain respectively. The two conserved domains in Chit-42 gene were GH-18 chitnase-like and Glyco-18 whereas LRR-RI domain was conserved in gene I-7 (Fig. 4). According to THHM Model, trans-membrane helix was present in extending residues from 907 to 929 of I-7 gene (Fig. 5A), while it was absent in rest of four genes (B-glucanase, chit-42, I-2-2942 and NBS-15). The trans-membrane is represented by blue colour while signal peptide by red colour (Fig. 5B).

Multiple Sequence Alignment and Phylogenetic Analysis

Protein sequences of *Fusarium* resistance genes were used to find out the homologous sequences which served as a benchmark for developing resistance genes for unknown *Fusarium* related diseases in annual, biennial and perennial horticultural crops. The multiple sequence alignment showed the conserved region in *Fusarium* resistance genes (Fig. 6). The phylogenetic tree was divided into three main groups, namely A, B, and C (Fig. 7). The group A consist of one gene (I2 clone 2956), whereas group B had six genes (I2 clone 3333, I2 clone 2958, I2 clone 2950, I2 clone 3324, I2 clone 3336 and I2 clone 2942) belonging to dicot species, *Solanum lycopersicum*. The group C had highest (58 gene) *Fusarium* resistance genes which were further classified into three different subgroups *i.e.* C1, C2, and C3 on the basis of their protein motifs. The subgroup-C1 comprises of 19 genes (tomato

species) with four conserved motifs (motif-1, 2, 3 and 4) and subgroup-C2, a very complicated subgroup, comprising of 25 genes of 9 different plant species (*Arabidopsis*, banana, *Lilium*, melon, onion, rice, soya bean, tobacco and tomato) and one from yeast (*Saccharomyces cerevisiae*) and one from fungi (*Trichoderma harzianum*) with no conserved motif, whereas subgroup-C3 consisted of 13 genes belonging to banana with three conserved motifs (motif 1, 3 and 5).

The *Fusarium* resistance genes presented in group A, showed more similar conserved motif pattern with group B followed by subgroup C1 and C3 (Fig. 7). Subgroup C2 was unique as compared to other groups and subgroups.

Motif prediction

In-silico motif analysis of 65 *Fusarium* resistance genes revealed that out of total 65 sequences, only 46 sequences had a p-value less than 0.0001 and E-value less than 10.Overall five significant motifs with minimum and maximum width of 12 and 60 respectively were mined for resistance genes and were designated as motif 1, motif 2, motif 3, motif 4 and motif 5 (Table 2).

Among all the genes studied, motif-3, the most frequently occurred motif was present in 43 genes followed by motif-1 in 40 genes, motif-2 in 28 genes, motif-4 in 26 genes and motif-5 in 13 genes (Fig. S1).

The gene NBS-15, chit-42, Lr14-3-3 and beta-glucanase had one motif, NBS-16 gene had two motifs, whereas 16 genes had three motifs and 24 genes had four motifs (Fig. S1).

Cis-regulatory elements

Analysis of 65 resistance genes resulted in 92types of *cis*-regulatory elements. Among them CAAT-box had maximum (65) frequency (present in all the resistance genes) followed by G-box (54), Skn-1_motif (44), ARE (42), TC-rich repeats (36), CGTCA-motif (36), TGACG-motif (36), circadian (35), Sp1(35), O2-site (34), A-box (33), Box I (33), AAGAA-motif (32), MBS (32), TATA-box (32) and TATCCAT/C-motif (32) (Fig. S2). The functions of different *cis*-regulatory elements are described in Table 3. According to differences in function, *cis*-regulatory elements were divided into 11 categories along with unknown function category.

Fusarium resistance genes are highly polymorphic and have wide recognition specificities due to gene duplication, genomic forces (insertion, deletion etc.) and constant selection pressure by pathogen evolution. This leads to slow progress in development of *Fusarium* disease resistant varieties.

Protein information of *Fusarium* resistance genes can be used in identification of homologs in different plant genomes, preservation of the genetic code, trans-membrane alpha-helices of membrane proteins and also to measure the hydrophobicity of a specific peptide/protein in the sample.

Our study has provided valuable insights regarding multiple sequence alignment analysis, construction and visualisation of phylogenetic trees, motif prediction and physiochemical properties of protein sequences.

Out of 65 *Fusarium* resistance proteins, 33 protein sequences were stable in nature with pIof protein value ranging from 4.17 to 11.12.Protein pI is having an important role in finding the pH dependent characteristics of a protein (Talley and Alexov, 2010). The instability index of the protein sequences describes whether the protein is stable or not.

Table.1 List of various *Fusarium* resistance genes and their physico-chemical properties

Fusarium species	Gene source	Gene	GenBankID	Protein length	Mol. Wt.	pI.	I.I.	GRAVY
Fusarium oxysporum	Petunia	PhDef-1	AF507975.1	103	11361.4	6.54	49.77	0.265
	Petunia	PhDef1-(2)	HQ694498.1	103	11349.3	6.54	47.90	0.215
	Petunia	PhDef-2	AF507976.1	101	11049.0	5.16	50.94	0.176
	Lilium	Lr14-3-3	KF362120.1	259	29301.0	4.79	49.57	-0.459
F. oxysporum f. sp. *vasinfectum*	Cotton	NaD-1	AF509566.1	105	11721.9	6.56	61.25	0.137
Fusarium oxysporum f. sp. cubense	Banana	MusaBAG-1	KJ636053.1	152	16834.5	9.10	42.23	-0.375
	Banana	NBS-2	KF034946.1	174	19426.3	5.49	43.77	-0.236
	Banana	NBS-3	KF034947.1	174	19353.3	5.78	45.68	-0.286
	Banana	NBS-4	KF034948.1	174	19478.4	5.78	48.97	-0.268
	Banana	NBS-5	KF034949.1	174	19620.5	6.20	55.70	-0.351
	Banana	NBS-6	KF034950.1	174	19537.4	5.78	55.99	-0.294
	Banana	NBS-7	KF034951.1	174	19601.5	5.78	51.10	-0.326
	Banana	NBS-8	KF034952.1	174	19569.5	5.78	56.48	-0.307
	Banana	NBS-9	KF034953.1	174	19510.4	5.78	57.10	-0.279
	Banana	NBS-10	KF034954.1	174	19537.4	5.78	55.99	-0.294
	Banana	NBS-11	KF034955.1	174	19537.4	5.78	55.99	-0.294
	Banana	NBS-12	KF034956.1	174	19537.4	5.78	55.99	-0.294
	Banana	NBS-13	KF034957.1	174	19537.4	5.78	55.99	-0.294
	Banana	NBS-14	KF034958.1	174	19569.5	5.78	56.48	-0.307
	Banana	NBS-15	KF034959.1	178	20113.9	6.84	39.15	-0.458
	Banana	NBS-16	KF034960.1	212	24338.0	7.10	47.83	-0.507
	Yeast	CHS-7	NM-001179272.1	316	34898.2	5.46	23.31	0.898
Fusarium oxysporumf. sp. Cubense race 1	Onion	Ace-AMP-1	AF004946.1	132	15141.9	11.12	50.39	0.017
	Trichoderma	chit-42	S78423.1	423	46056.5	7.01	23.57	-0.311
	Soybean	beta1,3-glucanase	AF034106.1	246	26992.5	4.52	31.12	0.074
Fusarium oxysporum f. sp. *Cubenserace 4*	Rice	Tlp	XM-015763091.1	177	17992.1	5.07	43.10	0.185
Fusarium oxysporumf.sp. matthioli (FOM)	Arabidopsis	RFO-1	DQ023268.1	749	83435.4	6.43	44.85	0.200
Fusarium oxysporumf.sp. melonis	Melon	Fom-2-PI161375	AY619649.1	533	61579.9	8.45	49.13	-0.214
	Melon	Fom-2 (AnanasYokneum	AY619648.1	533	61611.8	8.04	47.99	-0.211
	Melon	Fom-2(Vedrantais)	AY619647.1	533	61611.8	8.04	47.99	-0.211
	Melon	Fom-2 (Durango)	AY619646.1	533	61611.8	8.04	47.99	-0.211
	Melon	Fom-2 protein gene (Partial cds)	AY619650.1	554	63811.1	7.29	45.64	-0.231
	Melon	Fom-2 protein gene (Complete cds)	DQ287965.1	1073	123571.8	7.13	43.42	-0.193
Fusarium oxysporum f. sp. *lycopersici*(Fol) races 1, 2 and	Tomato	I-2 (clone 3318)	DQ205979.1	242	27159.3	7.64	32.40	-0.291
	Tomato	I-2 (clone 2949)	DQ205977.1	242	27113.3	8.32	30.89	-0.240
	Tomato	I-2 (clone3325)	DQ205976.1	242	27113.3	8.32	30.89	-0.240
	Tomato	I-2 (clone-2950)	DQ205975.1	233	25999.9	6.86	28.72	-0.299
	Tomato	I-2 (clone-3324)	DQ205974.1	241	27017.0	6.43	33.51	-0.299
	Tomato	I-2 (clone-2942)	DQ205973.1	238	26643.7	7.64	28.50	-0.267
	Tomato	I-2 (clone-3333)	DQ205972.1	243	27315.5	7.64	32.31	-0.283
	Tomato	I-2 (clone-2957	DQ205971.1	243	27315.5	7.64	32.31	-0.283
	Tomato	I-2 (clone-3323)	DQ205970.1	243	27401.6	7.64	31.31	-0.300
	Tomato	I-2 (clone-3335)	DQ205969.1	243	27368.5	8.41	33.65	-0.312
	Tomato	I-2 (clone-2961)	DQ205968.1	243	26650.7	8.64	29.18	-0.239
	Tomato	I-2 clone-2956	DQ205967.1	243	27315.5	7.64	32.31	-0.283
	Tomato	I-2 (clone-2953)	DQ205966.1	242	27353.2	4.97	32.28	-0.331
	Tomato	I-2 (clone-2944)	DQ205965.1	230	26076.8	5.03	32.47	-0.330
	Tomato	I-2 (clone-3336)	DQ205964.1	243	27285.4	7.64	29.05	-0.281
	Tomato	I-2 (clone-2962)	DQ205963.1	242	27413.2	4.97	30.47	-0.347
	Tomato	I-2 (clone-2947)	DQ205962.1	236	26328.3	8.32	26.29	-0.240
	Tomato	I-2 (clone-2954)	DQ205961.1	242	27353.2	4.97	32.28	-0.331
	Tomato	I-2 (clone-3326)	DQ205959.1	243	27297.4	7.64	32.62	-0.272
	Tomato	I-2-(clone-2943)	DQ205958.1	242	27111.4	8.32	30.01	-0.258
	Tomato	I-2 (clone-3337)	DQ205957.1	243	27315.5	7.64	32.31	-0.283
	Tomato	I-2 (clone-2948)	DQ205956.1	242	27327.1	4.97	33.08	-0.350
	Tomato	I-2 (clone-3332)	DQ205955.1	243	27301.4	7.64	30.89	-0.284
	Tomato	I-2 (clone-2958)	DQ205954.1	239	26832.0	8.33	29.82	-0.295
	Tomato	I-2-Motelle	KR108299.1	1266	144801.0	5.87	49.43	-0.250
	Tomato	I-2 (Heamsona)	FJ843082.1	52	5710.4	4.17	88.31	0.146
	Tomato	I-2 (Complete cds)	AF118127.1	1266	144827.1	5.87	49.50	-0.252
	Tomato	I-3-SRLK-4	KP082942.1	831	91960.3	5.62	39.57	-0.145
	Tomato	I-3-SRLK-6	KP082944.1	834	93655.9	8.22	40.16	-0.185
	Tomato	I-3-SRLK-5	KP082943.1	841	94589.8	6.97	37.97	-0.177
Fusarium oxysporum f. sp. lycopersici (Fol) race 3	Tomato	I-7 gene	KT185194.1	966	108155.0	5.72	33.18	-0.097
	Tomato	I-7 (Partial)	KT185195.1	953	106545.2	5.49	33.85	-0.091

pI: Isoelectric Point, I.I.: Instability Index, GRAVY- Grand Average of Hydropathicity

Table.2 Details of five best motifs present in *Fusarium* resistance genes and their best possible matches

Motif	Motif width	Best possible match
Motif-1	39	CALMMGNEQMSMPCLSTEASWSLFKRHAMENMDPMGHPE
Motif-2	60	LSMDTKGKNLAAVPIVGMGGLGKTTLAKAAYNDERVQKHFVLKA WFCVSEVYDAFTITKG
Motif-3	39	QVKLKESLKGKKFVIVLDDVWNENYNEWNDLRNPFVQGD
Motif-4	39	LEEVGRQIAAKCKGLPLALKTLAGMLRPKSEIDEWKCIL
Motif-5	60	GKTTLLKTLNNELKENTRDYHVVIMIEVANSETLNVVDMQKIIANRL GLPWNESETERER

Table.3 Grouping of cis-regulatory elements in function categories

Functional categories	Type of *cis*-regulatory element
***cis*-acting regulatory element involved in light responsiveness**	G-box, Sp1, Box I, TCT-motif, GT1-motif, rbcS-CMA7a, CATT-motif, I-box, GATA-motif, Box 4, AAAC-motif, ACA motif, as-2-box, ATCT-motif, 4cl-CMA2b, ATCC-motif, CHS-CMA2a, CHS-CMA2a, GA-motif, 3-AF1 binding site, CAG-motif, CGT-motif, CHS-CMA1a, GAG-motif, Gap box, GTGGC-motif, LAMP-element, MRE, TCCC-motif, TGG-motif, AE-box, ACE, L-box, ATC-motif, MNF1
Hormone Responsive *cis*-acting regulatory	**Abscisic acid responsive:** ABRE, Motif IIb **Auxin-responsive:** TGA-element, AuxRRcore **Ethylene-responsive:** ERE **Gibberellin-responsive:** P-box, GARE-motif, TATC-box **Salicylic acid responsive:** TCA-element
***cis*-acting regulatory element involved in biotic stress**	**TC-rich repeats (**defense and stress responsiveness**), Box-W1** (Fungal elicitor), EIRE (Elicitor-responsive), **AT-richsequence** (Maximal elicitor-mediated activation), **CGTCA-motif** (MeJa-responsiveness), **TGACG-motif** (MeJa-responsiveness)
***cis*-acting regulatory element involved in abiotic stress**	**MBS (**drought-inducibility**), HSE** (heat stress), **ARE** (anaerobic induction), **GC-motif** (anoxic specific inducibility), **WUN-motif** (Wound-responsive) **LTR** (low-temperature responsiveness), **MBSI** (flavonoid biosynthetic genes regulation)
Common *cis*-acting element in promoter and enhancer regions	CAAT-box
Core promoter element around -30 of transcription start	TATA-box
***cis*-acting element conferring high transcription levels**	5UTR PY-rich stretch
***cis*-acting regulatory element involved in circadian control**	Circadian
Unknown Function	AAGAA-motif, G-box, W-box, AC-I, TCCACCT-motif, GCC box, Y-box, Box E, F-box, Unnamed_1, 2, 3, 4, 8, 11,13 and 16
Mediating transactivation by MYB transcription factors during lignin biosynthesis	AC-I
Plant Development Process	O_2-site (Zein metabolism regulation), RY-element (seed-specific regulation), Skn-1_motif (endosperm expression), GCN4_motif (endosperm expression), CCGTCC-box (meristem specific activation)
***cis*-acting regulatory element involved in binding**	**Protein binding:** CCAAT-box **Mybhv1 binding site:** CCAAT-box **DNA binding protein :** OBP1 site **AT-rich DNA binding protein (ATBP-1):** AT-rich element

Supplementary Table.1 Secondary structure analysis of five *Fusarium* resistance genes

Species	Gene	Secondary structure	Percentage
Tomato	I-2-clone-2942	Disordered	18 %
		Alpha helix	49 %
		Beta strand	10 %
		TM helix	0 %
	I-7	Disordered	27 %
		Alpha helix	20 %
		Beta strand	23 %
		TM helix	3 %
Banana	NBS-15	Disordered	16 %
		Alpha helix	52 %
		Beta strand	10 %
		TM helix	0 %
Trichodermaharzianum	Chit-42	Disordered	15 %
		Alpha helix	34 %
		Beta strand	18 %
		TM helix	0 %
Soybean (Glycine max)	Beta-glucanase	Disordered	17 %
		Alpha helix	40 %
		Beta strand	24 %
		TM helix	0 %

Supplementary Table.2 Ramachandran plot statistics of five different genes

Species	Genes	Number of residues Per cent		
		Favoured region	**Allowed region**	**Outlier region**
Tomato (*Solanumlycopersicon*)	I-2-clone-2942	84.7%	8.9%	16.4%
	I-7	76.7%	15.2%	8.1%
Banana (*Musa paradisiaca*)	NBS-15	85.2%	9.7%	5.1%
Trichodermaharzianum	Chit-42	97.4%	2.1%	0.5%
Soybean (*Glycine max*)	Beta-glucanase	94.2%	5.8%	0.0%

Supplementary Table.3 The 3D analysis and Z-score prediction of five *Fusarium* resistance genes

Species	Gene	Z-Score	3D-1D profile >= 0.2
Tomato (*Solanumlycopersicon*)	I-2-clone- 2942	-4.68	56.36 %
	I-7	-3.30	76.09 %
Banana (*Musa paradisiaca*)	NBS-15	-2.95	41. 57 %
Trichodermaharzianum	Chit-42	-7.66	89.60 %
Soybean (*Glycine max*)	Beta-glucanase	-7.35	95.08 %

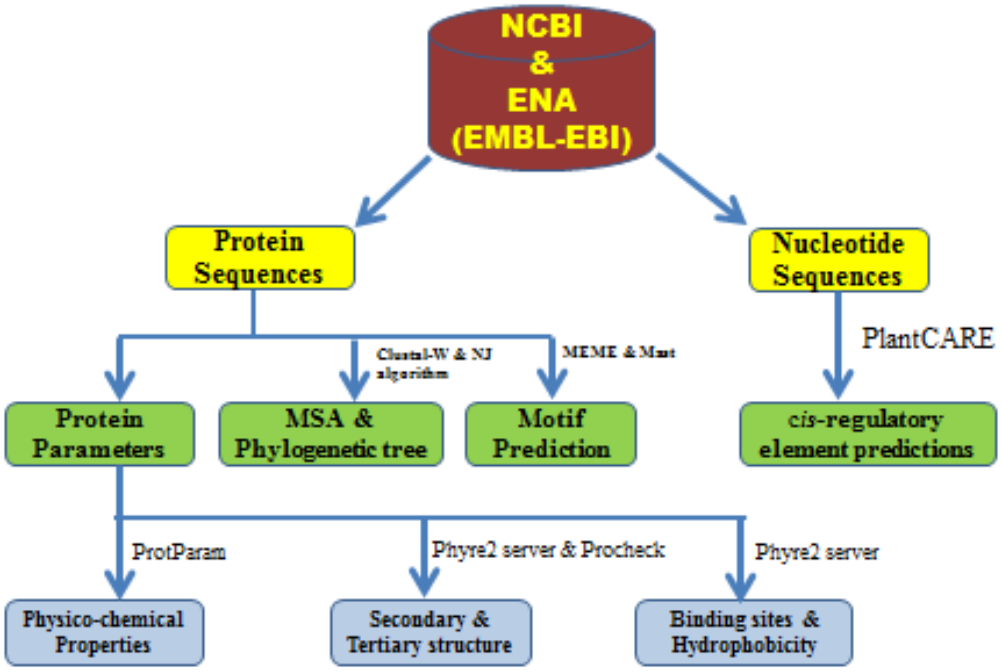

Fig. 1 : Flow chart of in-silico analysis of potential *Fusarium* resistance genes

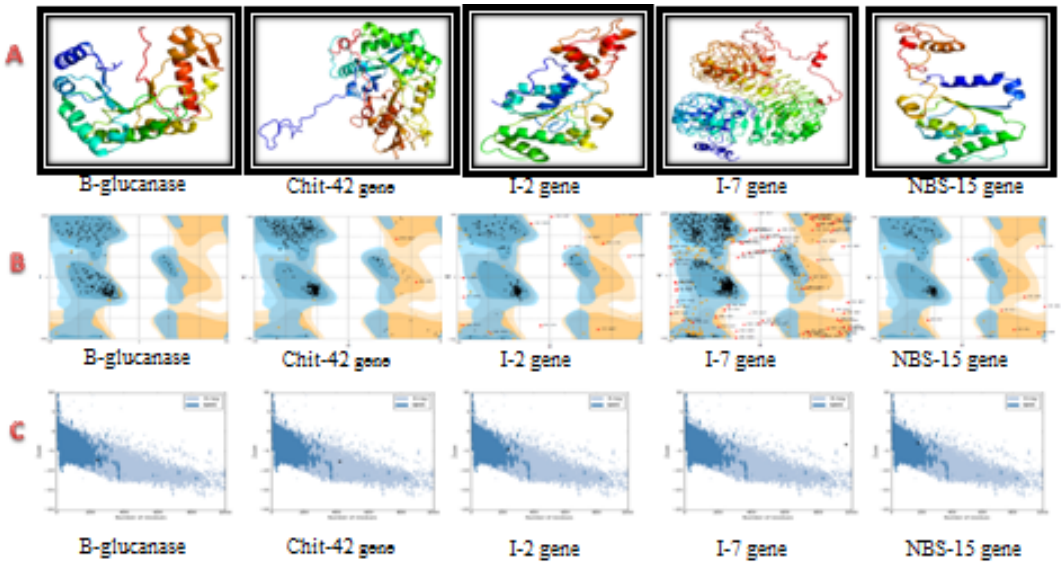

Fig. 2: A) Protein 3D structure predicted by Phyre2server of five *Fusarium* resistance genes B) Ramchandran plot of five *Fusarium* resistant gene analyzed through RAMPAGE and the residue are divided into three regions i) Favored region [■ ▲ , x] ii) Allowed region [■ ,▲, x] and iii) outlier region (white in colour). C) ProSA-web z-scores of five proteins of *Fusarium* resistance genes (B-glucanase, chit-42, I-2-2942, I-7, and NBS-15)

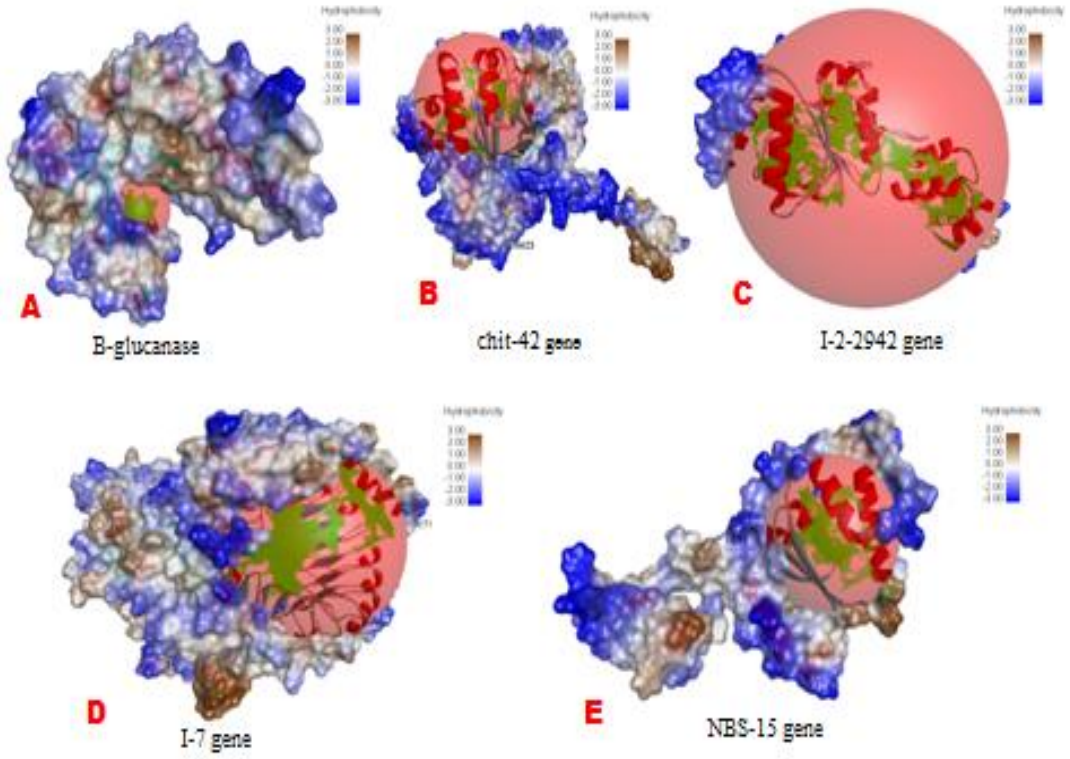

Fig. 3: Binding sites and hydrophobicity analysis in five stable protein sequences of *Fusarium* resistance genes : A,) B-glucanase, B) chit-42 gene, C) I-2-2942, D) I-7 gene and E) NBS-15

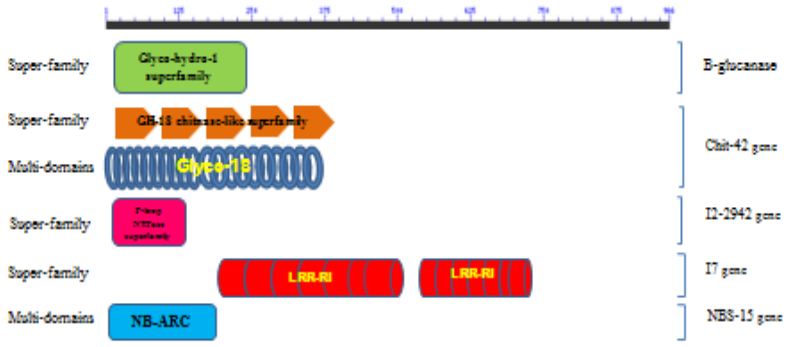

Fig.4: Domain architecture variations in five *Fusarium* resistance genes. The conserved protein domains were retrieved by using the Conserved Domains Database (CDD) database of NCBI.

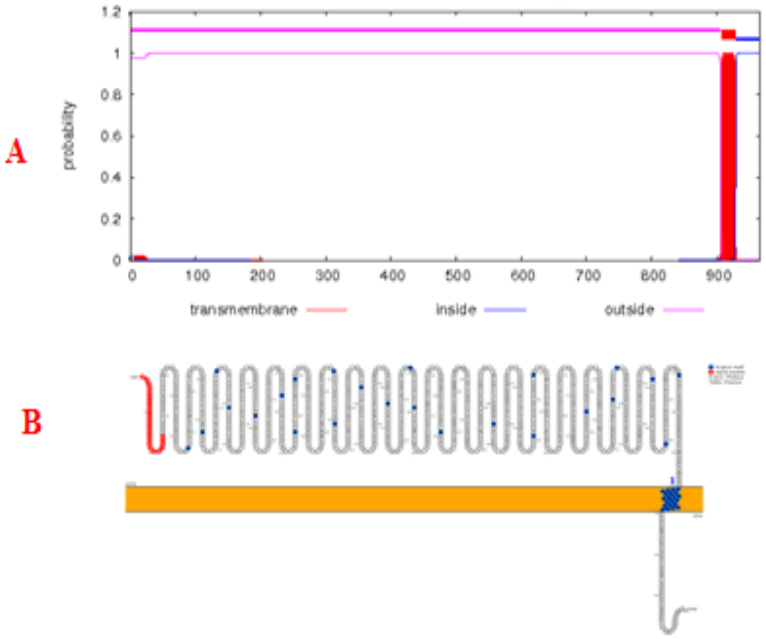

Fig.5: A) Trans-membrane prediction through Trans-membrane Helices Hidden Markov Model (TMHMM Server). B) Visualization of the trans-membrane structure drawn through the plotter tool.

Fig. 6Multiple sequence alignment and sequence comparison of 65 *Fusarium* resistance genes

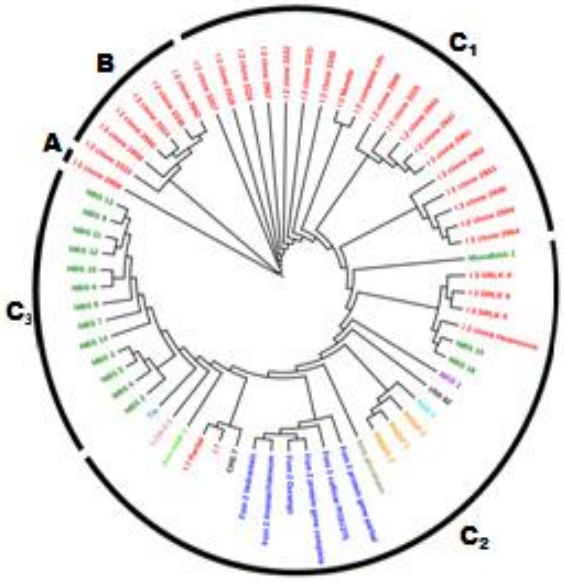

Fig. 7-Neighbor-joining phylogenetic tree based on the ClustalW alignment of *Fusarium* resistance genes.

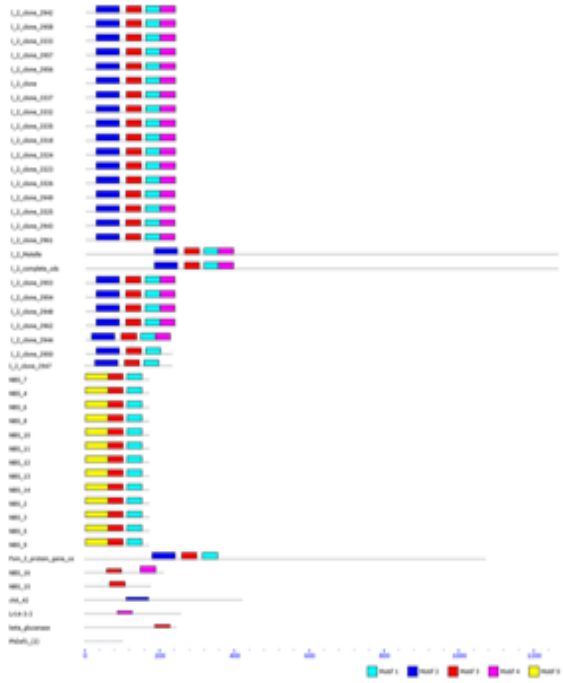

Fig. S1: Protein motif analysis of *Fusarium* resistance genes having p-value less than 0.0001

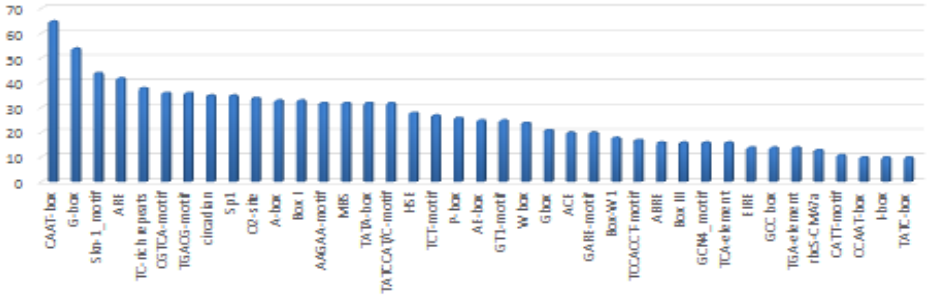

Fig. S2. C*is*-regulatory elements (CRE's) having >10 repetition among *Fusarium* resistance genes

Our study revealed that GRAVY score varied from -0.507 (NBS-16) to 0.898 (CHS-7). Similar reports were made earlier by other workers (Zhao and London, 2006; Crasto, 2010; Mallikarjuna *et al.,* 2016).

The alpha helix, beta strand and disordered percentage of secondary structure varied from 20 to 52 %, 10-24% and 15 to 27 % respectively. This information of secondary structure of protein will be helpful in understanding both the mechanisms of folding and the biological activity of proteins (Sivan *et al.,* 2007).The 3D structure predicted the arrangement of secondary structures as well as their side chains into three-dimensional space. The Z score of five stable *Fusarium* resistance genes (I-2, I-7, NBS-15, chit-42 and B-glucanase) were -7.35, -7.66, -4.68, -3.30 and -2.95 respectively. The negative value of Z-score indicated that these structures are reliable. Similar negative value of Z score was estimated in AC1 proteins of begomovirus strains (Prajapat *et al.,* 2007) and chitinase gene family in wheat (Mishra *et al.,* 2015).

Protein binding site is region in a protein on which specific molecules/ions attach to form a chemical bond. The knowledge of hydrophobicity will be helpful to have an idea regarding protein folding, prediction of trans-membrane sequences and water-protein-lipid interaction. In the current study, the protein binding sites and hydrophobicity level observed in beta-glucanase, I-2-clone-2942, I-7, chit-42 and NBS-15 suggested that the distribution pattern in these five genes for binding site is more, while hydrophobicity is very low to low. In proteins, binding sites are more active sites which are mainly due to hydrophobic nature and also play an important role in the interaction specificity (Labute and Santavy, 2007; Stank *et al.,* 2016). It is necessary to understand the distribution pattern of hydrophobicity in protein sequences and structures because hydrophobicity plays an important role in development of protein structures (Sandelin, 2004). The information on protein binding sites and receptor functionalities in horticultural crops is still lacking. Our findings on hydrophobicity of protein sequences show potential to find the crystallographic determinations of each binding sites and offers possibilities for developing varieties resistant to *Fusarium* diseases in different crops. Protein domain dynamics play crucial roles for molecular recognition and signalling processes in cells.

The GH18 (Glycoside Hydrolase18) domain was found in B-glucanase gene, while NB-ARC domain was found in NBS-15 gene. GH18, a multigene family of chitinases plays various roles in embryonic development and

allergic inflammation (Huang *et al.,* 2012). In the disease resistance genes or proteins, the NB-ARC acts as molecular switch to help the plant against biotic and abiotic stresses (Pal *et al.,* 2007; Van *et al.,* 2008). In our results, two of the genes (chit-42 and NBS-15) had multi-domain proteins. These multiple domain proteins are highly stable and have benefits in terms of folding compared to single domain proteins in the cell (Bhaskara *et al.,* 2011). The conserved domains namely GH18, NB-ARC, LRR-RI,P-loop NTPase superfamily andGlyco-18 present in resistance genes are independently stable sequences which can be used to develop chimeric proteins through genetic engineering for crop improvement. In our results, trans-membrane helix was present in I-7 gene; while it was absent in rest of four genes. Trans-membrane proteins in the resistance genes regulate the information and other different substances from inside and outside of the cell and are actively involved in a various other biological processes (Reynolds *et al.,* 2008).

Multiple sequence alignment (MSA) is becoming powerful tool in plant science for estimating phylogenetic trees, illuminating functionally important regions, structure prediction of proteins (secondary and tertiary structure), protein function, structure of RNAs (Kemena and Notredame, 2009), biological function analysis and performing the task of next-generation sequencing (Ortuno *et al.,* 2013). In the present study, protein sequences of 65 *Fusarium* resistance genes were used to find out the homologous sequences and multiple sequence alignment, and the conserved regions obtained will be helpful in knowing the binding sites, active sites and development of molecular makers.

In-silico motif analysis of 65 *Fusarium* resistance genes revealed that motif-3 was present in 43 genes followed by motif-1 in 40 genes. The gene NBS-15, chit-42, Lr14-3-3 and beta-glucanase had one motif; NBS-16 gene had two motifs. The statistical significance of motif prediction was correlated with biological significance, indicating a valuable reason for the motifs analysis as also reported earlier by Zhang *et al.,* (2009). Motifs information can be used for developing resistance genes and makers, to perform clustering (Broin *et al.,* 2015), gene expression analysis study (Jensen *et al.,* 2005, Huber and Bulyk, 2006) and discovery of homology relations (Stewart, 2016), family classification (Blekas *et al.,* 2005; Eser *et al.,* 2013), discovery of sub-families in large protein families (Leonardi and Galves, 2005) and new signalling pathways (Ma *et al.,* 2013).

Phylogenetic tree results were in agreement with the results of conserved motif analysis. The phylogenetic analysis of *Fusarium* resistance genes presented in group A showed more similar conserved motif pattern with group B, followed by subgroup C1 and C3 (Fig. 7). Evolutionary forces probably caused these changes during evolutionary processes. Subgroup C2 was unique as compared to other groups and subgroups which could be due to some genomic forces (insertion, deletion etc.) which might affect the different *Fusarium* resistance gene structures.

We observed various *cis*-regulatory elements *i.e.* TC-rich repeats, Box-W1, EIRE, AT-rich sequence, CGTCA-motif; TGACG-motif which involved in crucial roles against *Fusarium* induced diseases in various horticultural crops. The spatial and temporal differences in gene expression levels of an organism are the result of interactions between DNA, RNA and protein. DNA *cis*-regulatory elements consist of promoters, enhancers and insulators which play important role in gene regulation process (Fiedler and Rehmsmeier, 2006; Ibraheem *et al.,* 2010; Himani *et al.,* 2014). To understand

the gene regulation process and the development process of organisms, the knowledge of *cis*-regulatory elements is indispensable. Our studies showed that Box-W1 *cis*-regulatory element is *Fusarium* elicitor responsive and provides insights in the evolution of *Fusarium* resistance genes.

The identification of these CRE and their CREs across species will allow an understanding of the diversity, distribution and evolutionary relationships of these genes (Wittkopp, and Kalay, 2012; Lemmon *et al.*, 2014; Hernandez-Garcia, and Finer, 2014).

In-silico analysis revealed the homology of 65 *Fusarium* resistance genes. Resistance genes can be used to improve plant resistance through genome editing and gene silencing strategies. The physicochemical properties of *Fusarium* resistance genes can be used for identification of homologs in different plant genomes.

Knowledge of phylogenetic trees, *cis*-acting elements and motif prediction generated from this study will give better understanding of the 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 5 stable *Fusarium* genes can be effectively used for *in silico* docking study for development of potential ligand molecules against *Fusarium* infection. Our study also identified potential candidate genes. The marker developed from the candidate genes will be useful to develop highly resistance genotypes.

Acknowledgement

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

References

Agrawal, A.A., 2011. Current trends in the evolutionary ecology of plant defence. Funct. Ecol. 25:420-32.

Agrawal, A.A., Fishbein, M., 2008. Phylogenetic escalation and decline of plant defense strategies. Proc. Natl. Acad. Sci. U.S.A..105:10057-60.

Agrawal, A.A., Fishbein, M., Halitschke, R., Hastings, A.P., Rabosky DL, Rasmann S., 2009. Evidence for adaptive radiation from a phylogenetic study of plant defenses. Proc. Natl. Acad. Sci. U.S.A. 106:18067-72.

Azam, S., Munir, A., Khan, M.S., Fazal, S., Mehmood, A., 2017. In-Silico Identification of Novel Resistant Genes for Fungal Pathogen *Fusarium oxysporum* f. sp. cubense Race 4: Causative Agent of Banana Vascular Wilt Disease. J. Plant Biochem. Physiol. Pp. 5-2.

Bentley, A.R., Cromey, M.G., Farrokhi-Nejad, R., Leslie, J.F., Summerell, B.A., Burgess, L.W., 2006. *Fusarium* crown and root rot pathogens associated with wheat and grass stem bases on the South Island of New Zealand. Australas. Plant Pathol. 35: 495-502.

Bhaskara, R.M., Srinivasan, N., 2011. Stability of domain structures in multi-domain proteins. Sci. Rep.1:40.

Bhati, J., Chaduvula, K.P., Rai, A., Gaikwad, K., Soma, M., S., 2016. In-Silico Prediction and Functional Analysis of Salt Stress Responsive Genes in Rice (*Oryza sativa*). J. Rice Res. 4: 2.

Blekas, K., Fotiadis, D.I., Likas, A., 2005. Motif-based protein sequence classification using neural networks. J. Comput Biol. 12:64-82.

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

Broin, P.Ó., Smith, T.J., Golden, A.A., 2015. Alignment-free clustering of transcription

factor binding motifs using a genetic-k-medoids approach. BMC bioinformatics.16:22.

Chakraborty, S., Newton, A.C., 2011. Climate change, plant diseases and food security: an overview. Plant Pathol. 60:2-14.

Crasto, C.J., 2010. Hydrophobicity profiles in G protein-coupled receptor transmembrane helical domains. J Receptor Ligand Channel Res..123.

Dutt, S., Singh, V.K., Marla, S.S., Kumar, A., 2010. In silico analysis of sequential, structural and functional diversity of wheat cystatins and its implication in plant defense. Genomics, proteomics & bioinformatics, 8: 42-56.

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

Fiedler, T., Rehmsmeier, M., 2006. jPREdictor: a versatile tool for the prediction of cis-regulatory elements. Nucleic Acids Res..34:546-550.

Gao, X., Song, P.X., 2010. Composite likelihood Bayesian information criteria for model selection in high-dimensional data.J. Amer. Statist. Assoc. 105:1531-1540.

Garcia-Moreno, B. 2009. Adaptations of proteins to cellular and subcellular pH.J. Biol.8:98.

Gaucher, E.A., Kratzer, J.T., Randall, R.N., 2010. Deep phylogeny how a tree can help characterize early life on Earth. Cold Spring Harb. Perspect. Biol.2:a002238.

Hernandez-Garcia, C.M, Finer, J.J., 2014. Identification and validation of promoters and cis-acting regulatory elements. Plant Sci.217:109-19.

Himani, S., Sonia, S., Sneh, N., Rekha, M., Indu, S., Ravish, C., 2014. Computational analysis of cis-acting regulatory elements in 5'regulatory regions of sucrose transporter gene families in wheat and Arabidopsis. Res. J. Biotechnol.9:4.

Hu, B., Jin, J., Guo, A.Y., Zhang, H., Luo, J., Gao, G., 2014. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 817.

Huang, Q.S., Xie, X.L., Liang, G., Gong, F., Wang, Y., Wei, X.Q., Wang, Q., Ji, Z.L., Chen, Q.X., 2012. The GH18 family of chitinases: their domain architectures, functions and evolutions. Glycobiology.22:23-34.

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

Ibraheem, O., Botha, C.E., Bradley, G., 2010. In silico analysis of cis-acting regulatory elements in 5′ regulatory regions of sucrose transporter gene families in rice (*Oryzasativa Japonica*) and *Arabidopsis thaliana*. Res. J.Biotechnol.34:268-83.

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

Johnson, L.S., Eddy, S.R., Portugaly, E., 2010. Hidden Markov model speed heuristic and iterative HMM search procedure. BMC bioinformatics.11:431.

Joshi, R.K., Nayak, S., 2013. Perspectives of genomic diversification and molecular recombination towards R-gene evolution in plants. Physiol. Mol. Biol. Plants.19:1-9.

Kemena, C., Notredame, C., 2009. Upcoming challenges for multiple sequence alignment methods in the high-throughput era. Bioinformatics.25:2455-65.

Krogh, A., Larsson, B., Von Heijne, G., Sonnhammer, E. L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J.Mol.Biol. 305: 567-580.

Labute, P., Santavy, M., 2007. Locating binding sites in protein structures. J. Chem. Comput. Group.

Lemmon, Z.H., Bukowski, R., Sun, Q., Doebley, J.F., 2014. The role of cis regulatory evolution in maize domestication. PLoS Genet.10:e1004745.

Leonardi, F., Galves, A., 2005. Sequence motif identification and protein family classification using probabilistic trees. In Brazilian Symposium on Bioinformatics.190-193. Springer Berlin Heidelberg.

Lescot, M., Dehais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouze, P., 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 Res.30:325-7.

Lovell, S.C., Davis, I,W., Arendall, W.B., de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S., Richardson, D.C.,2003. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins: Struct. Funct.Bioinf.50: 437-450.

Luthy, R., Bowie, J.U., Eisenberg, D., 1992. Assessment of protein models with three-dimensional profiles. Nature.356:83.

Ma, S., Shah, S., Bohnert, H.J., Snyder, M., 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., Gupta, H.S.,2016. *In-silico* characterisation and comparative mapping of yellow stripe like transporters in five grass species. Indian J. Agric. Sci.86:621-627

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I, Lanczycki, C.J., 2014.CDD: NCBI's conserved domain database. Nucleic Acids Res.222-226.

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

Nakashima, K., Ito, Y., Yamaguchi-Shinozaki, K., 2009. Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis and grasses. Plant Physiol.149:88-95.

Nelson, P.E., Dignani, M.C., Anaissie, E.J., 1994. Taxonomy, biology, and clinical aspects of Fusarium species. Clin. Microbiol.*Rev. 7*: 479-504.

Nottensteiner, M., 2015. Molecular Characterization of Retroelement Encoded ROPIP1 as Virulence Effector of *Blumeriagraminis f. sp. hordei* (Doctoral dissertation, Dissertation, Munchen, Technische Universitat Munchen..

Omasits, U., Ahrens, C.H., Müller, S., Wollscheid, B., 2013. Protter: interactive protein feature visualization and integration with experimental proteomic data. Bioinformatics.btt607.

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

Pal, A., Chakrabarti, A., Basak, J., 2007. New motifs within the NB-ARC domain of R proteins: Probable mechanisms of integration of geminiviral signatures within the host species of Fabaceae family and implications in conferring disease resistance. J. Theor. Biol.246:564-73.

Prajapat, R., Marwal, A., 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. J. Viruses.

Reynolds, S.M., Kall, L., Riffle, M.E., Bilmes, J.A., Noble, W.S., 2008. Transmembrane topology and signal peptide prediction using dynamic bayesian networks. PLoS Comput. Biol. 4:e1000213.

Ronsted, N., Symonds, M.R., Birkholm, T., Christensen, S.B., Meerow, A.W., Molander, M., Molgaard, P., Petersen, G., Rasmussen, N., Van, Staden, J., Stafford, G.I., 2012. Can phylogeny predict chemical diversity and potential medicinal activity of plants? A case study of Amaryllidaceae. BMC Evol. Biol.12:182

Sandelin, E., 2004. On hydrophobicity and conformational specificity in proteins. Biophys. J.86:23-30.

Saremi, H., Burgess, L.W., 2010. Effect of soil temperature on distribution and population dynamics of *Fusarium* species. J. Agric. Sci.Tech. 2:119-125.

Sivan, S., Filo, O., Siegelmann, H., 2007. Application of expert networks for predicting proteins secondary structure. Biomol. Eng.24:237-243

Stank, A., Kokh, D.B, Fuller JC, Wade RC. 2016. Protein Binding Pocket Dynamics. Acc. Chem. Res.49:809-15.

Stewart, Jr. C.N., 2016. Plant biotechnology and genetics: principles, techniques, and applications. John Wiley & Sons.432

Talley, K., Alexov, E., 2010. On the pH- optimum of activity and stability of proteins. Proteins: Struct. Funct.Bioinf.78:2699-706.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol.30:2725-9.

Van den, B. E., Hofberger, J.A., Schranz, M. E., 2016. Flower power and the mustard bomb: Comparative analysis of gene and genome duplications in glucosinolate biosynthetic pathway evolution in Cleomaceae and Brassicaceae. Am. J. Bot.103:1212-22.

Van, O.G., Mayr, G., Kasiem, M.M., Albrecht, M., Cornelissen, B.J., Takken, F.L., 2008. Structure-function analysis of the NB-ARC domain of plant disease resistance proteins. J. Exp. Bot.59:1383-97.

Werner, G.D, Cornwell, W.K, Sprent, J.I, Kattge, J, Kiers, E.T., 2014. A single evolutionary innovation drives the deep evolution of symbiotic N2-fixation in angiosperms. Nat. Commun.5

Wiederstein, M., Sippl, M.J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res.35:407-410.

Wittkopp, P.J., Kalay, G., 2012. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nature Rev. Genet.13:59-69.

Zhang, S., Su, W., Yang, J. 2009. ARCS-Motif: discovering correlated motifs from unaligned biological sequences. Bioinformatics.25:183-189.

Zhao, G., London, E., 2006. An amino acid "transmembrane tendency" scale that approaches the theoretical limit to accuracy for prediction of transmembrane helices: relationship to biological hydrophobicity. Prot. Sci.; 15: 1987-2001.

How to cite this article:

Ashok Yadav, Usha Kalidindi, Pawan Kumar Jayaswal, Shikha Mittal and Nepolean Thirunavukkarasu. 2018. Molecular Characterization and Evolutionary Analysis of Potential *Fusarium* Resistant Genes for Crop Improvement. *Int.J.Curr.Microbiol.App.Sci.* 7(02): 3052-3070. doi: https://doi.org/10.20546/ijcmas.2018.702.369