Journal of Plant Genomics
ISSN: 0976-8823 & E-ISSN: 0976-8831, Volume 2, Issue 1, 2012, pp.-18-27.
Available online at http://www.bioinfo.in/contents.php?id=78

GENOME WIDE ANALYSIS OF DISEASE RESISTANCE *MLO* GENE FAMILY IN SORGHUM [*Sorghum bicolor* (L.) MOENCH]

SINGH V.K.[1], SINGH A.K.[2], CHAND R.[3] AND SINGH B.D.[4]*

[1]Centre for Bioinformatics, Faculty of Science, Banaras Hindu University, Varanasi- 221 005, UP, India
[2]Department of Genetics and Plant Breeding, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi- 221 005, UP, India
[3]Department of Mycology and Plant Pathology, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi- 221 005, UP, India
[4]School of Biotechnology, Faculty of Science, Banaras Hindu University, Varanasi- 221 005, UP, India
*Corresponding Author: Email- brahmadsingh@gmail.com

Received: June 12, 2012; Accepted: July 18, 2012

Abstract- Powdery mildew of sorghum (*Erysiphe pisi* var. *pisi*) is of worldwide occurrence and causes substantial yield losses. As far as resistance against powdery mildew is concerned, the presence of powdery mildew-resistance gene O (Mlo) plays a key role. The *Mlo* gene was first identified in barley, and the Mlo protein was found to be an integral plasma membrane-localized protein that possesses seven transmembrane regions. In the present work a total of 12 well reported *Mlo* genes from *Oryza sativa,* 15 well characterized *Mlo* genes from *Arabidopsis thaliana* genome were taken for comparative studies in sorghum (*Sorghum bicolor*). Comparative analysis of Mlo proteins from *S. bicolor* genome revealed the presence of 13 hypothetical genes within the genome. Map viewer analysis indicates that the predicted *S. bicolor SbMlo* genes are distributed on eight of the ten chromosomes. Sorghum chromosome 9 has 3 genes; chromosome 10 and chromosome 1 have 2 genes each, while chromosomes 2, 3, 4 and 5 have 1 gene each. Sub-cellular localization of identified Mlo proteins encoded by genes *SbMlo1, SbMlo2, SbMlo4, SbMlo5, SbMlo6, SbMlo9, SbMlo10, SbMlo11, SbMlo12* and *SbMlo13* are present in plasma membrane; *SbMlo3* product is located in endoplasmic reticulum, *SbMlo7*encoded protein is located in vacuolar membrane and *SbMlo8* product is present in the nuclear region. *In silico* characterization (using phylogenetic classification, motif analysis and *cis*-acting elements studies) suggested its diverse function associated with disease resistance based on specific expression containing fungal elicitor responsive elements. Gene specific primers, expression primers and universal primer were designed to check the expression and availability of *SbMlo* genes through wet lab experimentation.
Keywords- *Mildew resistance, Sorghum bicolor*, comparative analysis, *SbMlo,* disease resistance, powdery mildew resistance

Citation: Singh V.K., et al. (2012) Genome Wide Analysis of Disease Resistance MLO Gene Family in Sorghum [*Sorghum bicolor* (L.) Moench]. Journal of Plant Genomics, ISSN: 0976-8823 & E-ISSN: 0976-8831, Volume 2, Issue 1, pp.-18-27.

Introduction

The powdery mildew disease in cereals is caused by an obligate biotrophic ascomycete fungus belonging to order Erysiphales. The *Mlo* (powdery-*mil*dew-resistance gene *o*) gene was first identified in barley [6] and it constituted the largest family of seven-transmembrane (7TM) domain proteins found in both monocot and dicot plants [12]. Many Mlo homologs have been identified in various plants; for example, thale cress (*Arabidopsis thaliana*) and rice (*Oryza sativa*), whose full genome sequences are available, contain 15 *AtMlo* [11] and 12 *OsMlo* [22] genes, respectively. Similarly, multiple *Mlo* genes have been identified in othe species, *e.g.,* 17 *VvMlos* in grapes [16], 7 *TaMlos* in wheat [21], 2 *BrMlos* in brassica [25], 2 *LjMlos* in lotus [25], 2 *CaMlo*s in capsicum [25], and 3 *LeMlo* in tomato [25]. *Mlo* gene family members are reported to play crucial roles in modulating defense responses and cell death [19,25,26,38].

Cereals and millets belong to family Poaceae and provide most of our calorie and protein requirements [35]. Disease resistance gene cluster synteny and genome wide gene family study can give a better understanding of the gene organization at genome

level, which may facilitate genetic improvement of crops. *Sorghum* is the fifth most important 'cereal' crop in the world based on total grain production [15]. *Sorghum* has a relatively small genome (750 Mbp) and it has diverged from maize and rice; this would greatly aid the discovery and analysis of disease resistance genes through comparative genomics. *Sorghum* is a C4 plant that exhibits drought tolerance, stalk reserve retention capacity and potential to produce more grain per unit photosynthetic area [24]. Expression of *Mlo* genes can be affected by biotic and abiotic stimuli and the inducibility of *Mlo* expression under a range of conditions suggested a broad role of *Mlo* genes [26]. Identification of *Mlo* genes in other crops would be helpful in understanding the evolution and functional divergence at this locus, and it would also contribute to breeding of powdery mildew resistant cereal varieties.

This paper reports genome wide *in silico* identification of putative *SbMlo* gene family of *S. bicolor* (L.) Moench using the recently available whole genome shotgun sequence of *Sorghum* for annotation, chromosomal localization, gene organization and phylogenetic tree inferences based on function motifs. Further, comparative phylogeny of *O. sativa, A. thaliana* and *Sorghum Mlo* gene families has been attempted, and the putative functions of the predicted *SbMlo* genes were investigated by analyzing the *cis*-regulatory elements and transcription factors associated with these genes in the promoter and transcribed regions.

Materials and methods

Database Search for the Identification of *MLO* Gene Family Members in *S. bicolor*

The *Arabidopsis* (*A. thaliana*) genome contains 15 *Mlo* genes [7] designated as *AtMlo*s [11]. The Rice (*O. sativa*) genome contains 12 genes encoding homologs of Mlo protein; these genes were designated as *OsMlos* [29]. Based on the information provided by Chen et al. (2006), the *Arabidopsis Mlo* sequences were retrieved from TAIR web site (http://www.arabidopsis.org/) and reported rice *Mlo*s genes were retrieved from NCBI database [27]. The retrieved sequences were subjected to homology search with the available sequence information at NCBI database using BLASTn, BLASTx and tBLASTx [1,2] tools. Upstream and downstream sequences of *Mlo* domain homologs were retrieved from the whole genome shotgun sequence of *S. bicolor* for fishing out putative *SbMlo* genes. The annotated sequences were further subjected to bioinformatics tools and software, namely, FGENESH [30] for prediction of full length genes with putative full CDS and protein sequences.

Mapping of *Sbmlo* Genes on *Sorghum* Chromosomes and Determination of Intron/Exon Boundaries

Each of the *SbMlo* genes was positioned on *Sorghum* chromosomes by the BLASTn search with NCBI genomes (chromosome) database. The structures of predicted *SbMlo* genes were determined using FGENESH server and intron/exon boundaries were manually identified.

Sub-Cellular Localization and Trans-membrane Helix Prediction

The putative Mlo protein sequences of *S. bicolor* were subjected to protein functional analysis using PFAM version 24.0 [17] and MOTIFSCAN [14] databases. These sequences were then submitted to PSHORT (http://wolfpsort.seq.cbrc.jp/) server for identification of localization signals. For identification of transmembrane helices, HMMTOP (http://www.enzim.hu/hmmtop/html/submit.html) was used for finding the topology of proteins. For amino acid composition and physico-chemical properties, PROTPARAM (http://expasy.org/tools/protparam.html) was used.

MLO Protein Alignment and Phylogenetic Analysis

Identified putative *Mlo*s from *S. bicolor* were used for phylogenetic classification of *Mlo*s from *O. sativa* and *A. thaliana*. The sequences from different species were aligned using Clustalw [33] and phylogenetic inferences were constructed using MEGA5.0 [32].

Identification of Conserved Motifs/Transcription Factors and Cis-regulatory Elements

All Mlo proteins from *Arabidopsis*, rice and *Sorghum* were used for conserved motif study using MEME version 4.4.0 [4,5] in Mlo protein domain (Functional Signature Sequences) part. *Mlo* protein sequences were subjected to INTERPROSCAN version 4.4 [28] for protein functional analysis.

Designing of Molecular Markers (Primers) for Full Length Amplification and Gene Expression Study

Identified complete *Mlo* genomic sequences from *S. bicolor* were used to design primers for complete gene amplification. Expression primers to study the expression of identified genes during pathogenesis were designed using primer3 [31] based on hypothetical mRNA obtained from gene prediction tool.

Results

Domain based homology search using tBLASTn txid4558 of Whole-Genome-Shotgun Sequences (WGS) database identified 13 genes, designated as *SbMlo1* to *SbMlo13*, out of 14 hits form *S. bicolor* genome (Table 1).

Table 1- Genome wide identification of Mlo gene family in S. bicolor

Sl. No.	Gene Name	Primary Accession No.	Start	End
1	*SbMlo1*	ABXC01000744.1	818	6012
2	*SbMlo2*	ABXC01000489.1	131090	137248
3	*SbMlo3*	ABXC01000917.1	326708	330902
4	*SbMlo4*	ABXC01002307.1	95611	100790
5	*SbMlo5*	ABXC01002806.1	40722	44916
6	*SbMlo6*	ABXC01003117.1	79694	86858
7	*SbMlo7*	ABXC01004301.1	16063	22266
8	*SbMlo8*	ABXC01004476.1	55604	62500
9	*SbMlo9*	ABXC01006196.1	1	5000
10	*SbMlo10*	ABXC01006197.1	1	4531
11	*SbMlo11*	ABXC01006196.1	14548	21790
12	*SbMlo12*	ABXC01006822.1	44376	53573
13	*SbMlo13*	ABXC01006924.1	10660	15860

Chromosomal Localization of Predicted *SbMlo* Genes

Using gene prediction tool, 12 full length genes were successfully predicted out of the 13 identified genes, and their organization is summarized in Table 2. Chromosome 9 has three of these genes, chromosomes 1, 6 and 10 have two genes each, while chromosome 2, 3, 4 and 5 have one gene each. The number of exons per gene ranged from 8 to 15, but most of the genes had 11-14 exons. Thus, *Mlo* genes were distributed on eight out of the 10 chromosomes of *S. bicolor.*

Table 2- Organization of the S. bicolor Mlo genes.

Gene name	Chromosome No.	Exons	Introns (NC)
SbMlo1	1	12	11
SbMlo2	1	15	14
SbMlo3	2	12	11
SbMlo4	3	11	10
SbMlo5	4	8	7
SbMlo6	5	14	13
SbMlo7	6	14	13
SbMlo8	6	13	12
SbMlo9	9	14	13
SbMlo10	9	11	10
SbMlo11	9	14	13
SbMlo12	10	13	12
SbMlo13	10	9	8

Phylogenetic Relationships among *Mlo* Proteins of *Sorghum*

The complete catalog of Mlo proteins in a single plant species is useful for viewing the existing structural and functional diversity associated with their diverse role in plants. The evolutionary relationships among the 13 SbMlo proteins were analyzed by subjecting the amino acid sequences deduced from the identified 13 *SbMlo* genes for multiple sequence alignment. These proteins formed two main groups: Sb*Mlo*6 and Sb*Mlo*8 formed one group (group B), and the remaining 11Sb*Mlo*s formed the group A, which consisted of two subgroups (subgroups I and II had 8 and 3 proteins, respectively).

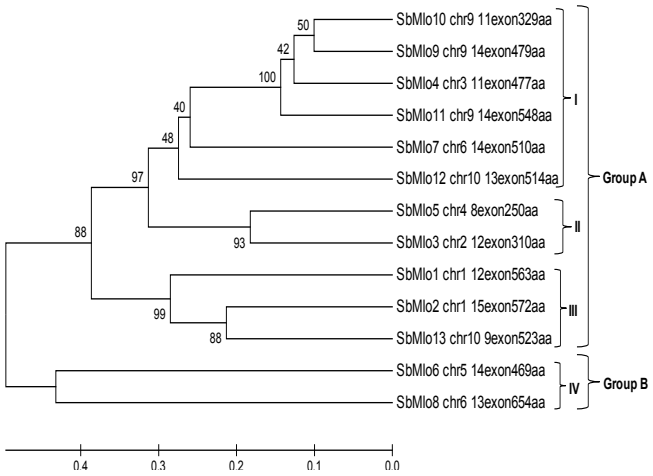

Fig. 1a- Phylogenetic classification of the13 identified Sb*Mlo* proteins from *S. bicolor*

The putative amino acid sequences of the 13 Sb*Mlo* were subjected to multiple sequence alignment analysis using clustalW. Four motifs, having consensus sequence and FWF residues, are totally conserved among the 13 Sb*Mlo* proteins (Fig. 1b). Evolutionary study suggests that Sb*Mlo*10 has 80.2% and 73.7% similarity with Sb*Mlo*9 and Sb*Mlo*11, respectively, Sb*Mlo*4 is 65.1% similar to Sb*Mlo*9, Sb*Mlo*3 has 62.3% identity with Sb*Mlo*5, and Sb*Mlo*2 and Sb*Mlo*1 are 58.8% and 50.9%, respectively, identical with Sb*Mlo*13. Similarly, Sb*Mlo*7 is 50.1 and 50.2% similar to Sb*Mlo*9 and Sb*Mlo*10, respectively, Sb*Mlo*12 is 47.4% identical with Sb*Mlo*9, and Sb*Mlo*6 has 42.6% identity with Sb*Mlo*8. Over this entire alignment pattern it was found that Sb*Mlo*10, Sb*Mlo*9 and Sb*Mlo*11 showed a high degree of similarity with each other, while Sb*Mlo*6 showed the lowest percentage of homology.

Fig. 1b- Alignment of 13 Sb*Mlo* proteins identified from *S. bicolor*

Gene structure prediction

The organization of the predicted *SbMlo* genes in terms of intron/exon distribution pattern shown in Table 3 and Fig. 2a. The minimum genomic gene size was 2207 bp for *SbMlo5*, while the maximum size was 6733 bp for *SbMlo12*. In terms of proteins, the minimum size was 250 aa (28.55 kDa) for Sb*Mlo*5, and the maximum size of 654 aa (74.47 kDa) was for Sb*Mlo*8. Compostional study of the indentified Sb*Mlo* proteins revealed that in most of the, Leucine was the most preponderant amino acid, followed by Val, Ala, Ser and Ile (Fig. 2b).

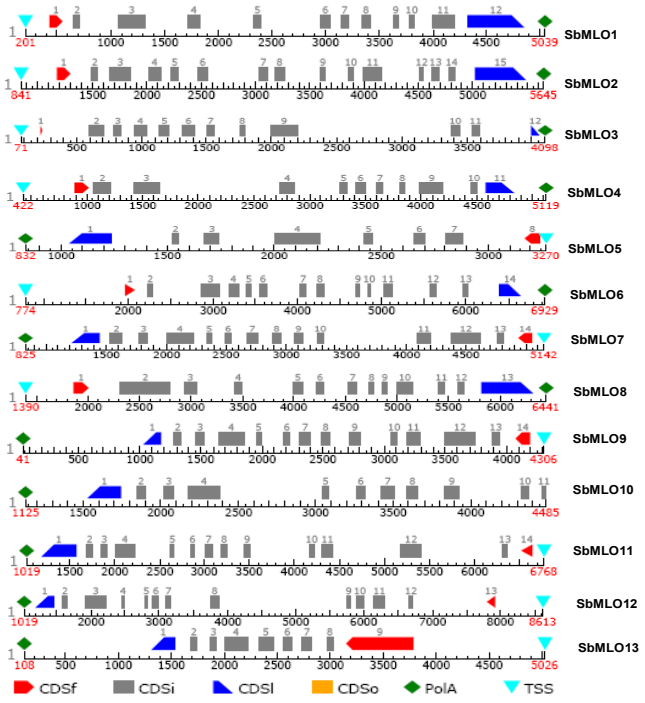

Fig. 2a- Complete gene organization including Exon-intron boundaries for each *SbMLO* genes

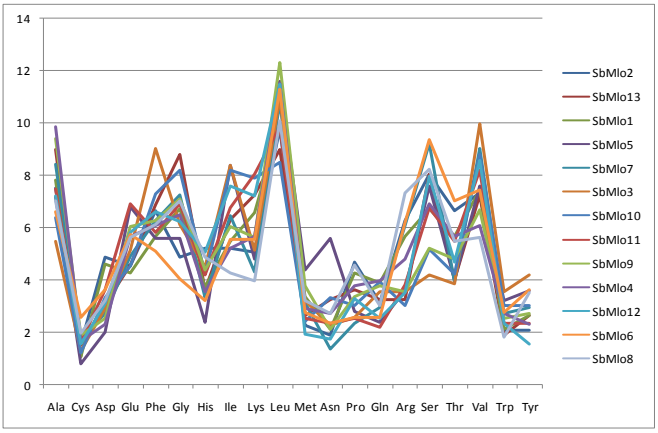

Fig. 2b- Amino acid composition comparartive graph of the 13 identifies Sb*Mlo* proteins

Table 3- SbMlo intron-exon boundaries, CDS size, full length of the encoded hypothetical amino acid sequence with molecular weight and theoretical pI.

Gene Name	Genomic size (introns+ exons, bp)	CDS size (bp)	Protein size (amino acid)	Molecular weight (Da)	Theoretical pI
SbMlo1	4422	1692	563	63599.9	9.53
SbMlo2	4314	1719	572	65051	9.13
SbMlo3	3839	993	310	36308.9	8.26
SbMlo4	3953	1434	477	53606.4	9.26
SbMlo5	2207	753	250	28554.4	6.6
SbMlo6	4695	1410	469	53872.5	9.01
SbMlo7	3836	1533	510	57328	9.35
SbMlo8	4463	1965	654	74471.6	9.21
SbMlo9	3171	1440	479	53972.3	8.27
SbMlo10	3008	990	329 (p.s)	---	---
SbMlo11	5461	1647	548	61792.9	8.73
SbMlo12	6733	1545	514	57780.7	8.99
SbMlo13	2477	1572	523	57725	9.3

Transmembrane helix prediction and toplology determination of the13 Sb*Mlo* proteins were done using HMMTOP [34]. The number of transmembrane helices ranged from 5 for Sb*Mlo*10 to 9 for Sb*Mlo*1 and Sb*Mlo* 9 (Table 4).

Table 4- Transmembrane helices prediction and toplology of the predicted 13 SbMlo proteins from S. bicolor.

Gene Name	N-terminus	Number of trans membrane helices	Positions of trans membrane helices	Total entropy of the model	Entropy of the best path
SbMlo1	OUT	9	15-32, 63-80, 160-179, 210-227, 232-251, 282-300, 305-323, 364-388, 407-425	17.0166	17.0204
SbMlo2	OUT	7	17-36, 63-81, 144-163, 265-284, 299-318, 349-372, 387-410	17.0175	17.0215
SbMlo3	IN	7	19-38, 97-114, 141-160, 165-182, 229-246, 265-284, 291-308	17.0165	17.0181
SbMlo4	IN	8	18-35, 58-75, 98-115, 227-244, 267-284, 289-306, 348-365, 370-387	17.0127	17.015
SbMlo5	IN	4	47-64, 69-86, 135-152, 165-182	17.0103	17.0115
SbMlo6	OUT	6	15-32, 63-80, 147-168, 207-224, 281-304,348-365	17.0173	17.0218
SbMlo7	OUT	7	13-30, 59-76, 154-173, 268-285, 292-309, 359-376, 395-416	17.0141	17.0178
SbMlo8	IN	8	24-41, 96-113, 132-149, 230-251, 299-316, 354-371, 378-395, 463-485	17.0214	17.0251
SbMlo9	OUT	9	14-31, 62-79, 106-125, 159-178, 240-257, 282-299, 306-323, 371-388, 407-426	17.0153	17.0196
SbMlo10	OUT	5	76-93, 110-129, 134-151, 191-215, 234-253	17.0172	17.0191
SbMlo11	OUT	7	17-34, 65-82, 158-180, 270-287, 294-311, 358-375, 394-416	17.0131	17.0161
SbMlo12	OUT	7	16-33, 64-83, 129-148, 265-284, 289-306, 363-382, 403-420	17.0102	17.0141
SbMlo13	IN	8	19-36, 67-84, 190-209, 251-267, 314-331, 338-355, 394-411, 442-464	17.0148	17.0197

Genome Wide Evolutionary Relationships Among *Sorghum*, Rice and *Arabidopsis Mlo* Gene Families

The predicted 13 SbMlo proteins were subjected to multiple sequence alignment along with the 15 *Arabidopsis* At*Mlo* and 12 *Oryza* Os*Mlo* proteins, and a phylogenetic tree was constructed by software ClustalX2.0.10 with UPGMA method and bootstrap analysis (1,000 reiterations) using MEGA5.0 (Fig. 3).

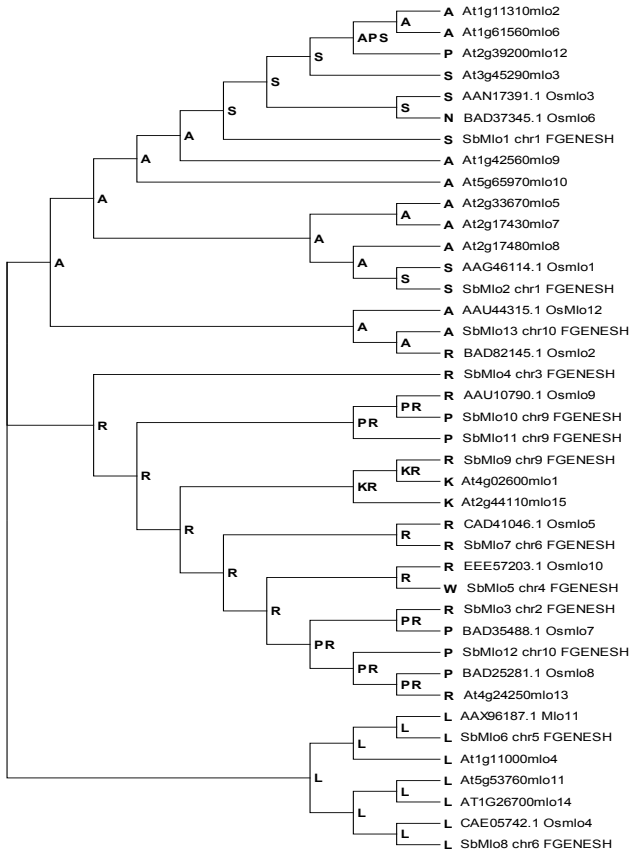

Fig. 3. Ancestral states of the phylogenetic tree were inferred using the maximum parsimony method; the total numbers of Mlo gene sequences taken were 15 from *A. thaliana*, 12 from *O. sativa* and 13 from *S. bicolor*.

The 40 *Mlo* proteins formed three groups: Group I is the largest with 17 Mlo proteins (9 *Arabidopsis*, 5 *Oryza* and 3 *Sorghum* proteins), group II is almost equally large with 16 proteins (3 *Ara-*

bidopsis, 5 *Oryza* and 8 *Sorghum* proteins), and group III is the smallest with only 7 proteins (3 *Arabidopsis,* 2 *Oryza* and 2 *Sorghum* proteins). Thus group I is dominated by At*Mlo*s, group II has predominantly Sb*mlo*s, while group III has almost equal representation from the three species. Further, group III is the most homogeneous of the three groups, while Group I and II show considerable sequence heterogeneity. The group I has At*Mlo* 2, 3, 5, 6,7, 8, 9, 10 and 12, Os*Mlo* 1, 2, 3, 6 and 12, and Sb*Mlo* 1, 2 and 13. The members of group II were At*mlo* 1, 13 and 15, Os*mlo* 5, 7, 8, 9 and 10, and Sb*mlo* 4, 5, 3, 7, 9, 10, 11and 12. The group III has At*mlo* 4, 11, and 14, Os*mlo* 4 and 11, and Sb*mlo* 6 and 8.

Motif and Transcription Factor Analysis

The 13 SbMlo proteins along with 15 At*Mlo* and 12 Os*Mlo* proteins were analyzed for the presence of conserved motifs using MEME software. A total of 26 conserved motifs were observed in the 40 proteins using 30 number of motifs with minimum width 50 and maximum width 100 parameter (Fig. 4a,b).

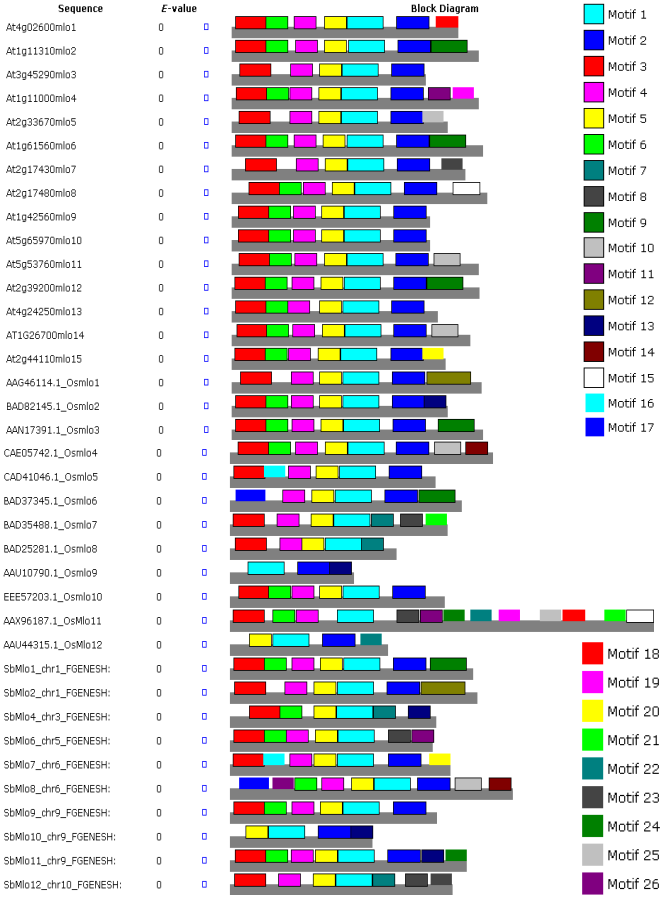

Fig.4a- Highly representative motif study using *Sorghum*, rice and *Arabidopsis Mlo* proteins.

Out of the 26 predicted motifs, Motif 01 was the most conserved pattern found in all the *Sorghum*, rice and *Arabidopsis Mlo* proteins. Motif 26 (in Sb*Mlo*13 and Sb*Mlo*8), Motif 24 (in Sb*Mlo*11 and Os*Mlo*11), Motif 23 (in Sb*Mlo*12 and At*Mlo*7), Motif 20 (in At*Mlo*15 and Sb*Mlo*7), Motif 17 (in Sb*Mlo*8 and Os*Mlo*6), Motif 16 (in Sb*Mlo*7 and Os*Mlo*5), Motif 14 (in Sb*Mlo*8 and Os*Mlo*4) and Motif 12 (in Sb*Mlo*2 and Os*Mlo*1) were the least conserved and were observed in only two *MLO* proteins each. Motif 11 was present in three proteins (Sb*Mlo*6, Os*Mlo*11 and At*Mlo*4), Motif 10 was detected in four *MLO* proteins (Sb*Mlo*8, Os*Mlo*4, At*Mlo*14 and At*Mlo*11), Motif 13 was found in five proteins (Sb*Mlo*11, Sb*Mlo*10, Sb*Mlo*4, Os*Mlo*9 and Os*Mlo*2), Motif 09 (Sb*Mlo*1, Os*Mlo*6, Os*Mlo*3, At*Mlo*12, At*Mlo*6 and At*Mlo*2) and Motif 08 (Os*mlo*7, Os*Mlo*11, Sb*Mlo*5, Sb*Mlo*6, Sb*Mlo*8 and Sb*Mlo*12) occurred in six *Mlo* proteins each, while Motif 07 was present in seven proteins (Os*Mlo*11, Sb*Mlo*8, Sb*Mlo*4, Sb*Mlo*5, Os*Mlo*7, Sb*Mlo*12 and Os*Mlo*8). Thus Motif 01 was present in all the 40 *Mlo* proteins, and Motifs07, 08 and 09 were the next most frequent (found in 7 and 6 *Mlo* proteins); the remaining motifs occurred in 5 or less number of *Mlo* proteins.

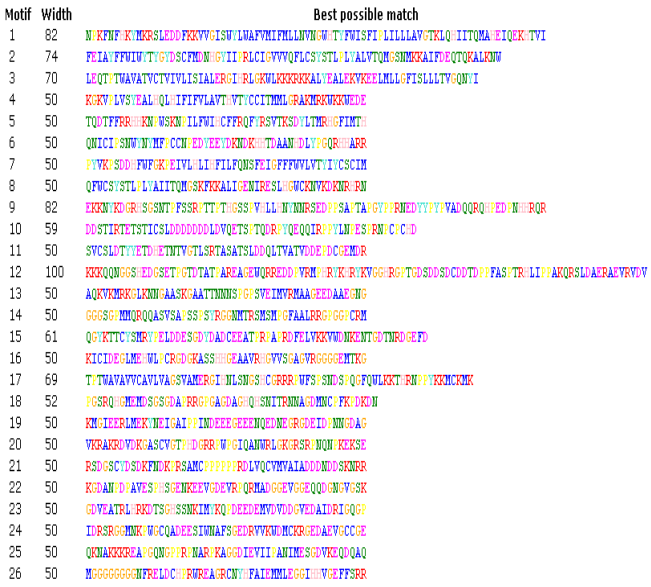

Fig.4b.- Multilevel consensus sequences for the MEME defined motifs observed among *Mlo* proteins from *Sorghum*, rice and *Arabidopsis.*

***Cis*-Regulatory Element Analysis**

The diverse functions attributed to different *Mlo* genes involve interactions of transcription factors with different conserved sequences in the promoter regions of the respective genes. *Cis-regulatory* element analysis was carried out with a view to understand the regulatory aspects of the different *SbMLO* genes. This was achieved by retrieving 500bp upstream sequences from the initiation codon of 12 putative *SbMlo* genes since the transcription start site (TSS) and initiation codon region are missing from the *SbMlo10* sequence due to non-completion of sequence. TATA and CAAT boxes were frequently located within -1 to -200bp. G-Box was the most frequent of the *cis*-elements; it was found in all the *Mlo* genes, except *SbMlo5* and *SbMlo10*. ABRE was observed in five genes (*SbMlo1, SbMlo3, SbMlo7, SbMlo9*, *SbMlo12*), A-box was found in seven genes (*SbMlo2, SbMlo4, SbMlo6, SbMlo8, SbMlo9, SbMlo11*, *SbMlo13*), ARE was present in four *SbMLO* genes (*SbMlo1, SbMlo7, SbMlo8,SbMlo9*), ACE in five genes (*SbMlo4, SbMlo5, SbMlo6, SbMlo9*, *SbMlo11*); AAGAA in two genes (*SbMlo5*, *SbMlo8*), AT-rich sequence in two genes (*SbMlo6, SbMlo7*), Box-3 and 4 in two genes (*SbMlo2*, *SbMlo7*), CCAAT in five genes (*SbMlo1, SbMlo7, SbMlo8, SbMlo11*, *SbM-*

lo12), CGTCA in six genes (*SbMlo1, SbMlo4, SbMlo5, SbMlo8,* Sb*Mlo*10, *SbMlo13*), CCGTCC in six genes (*SbMlo2, SbMlo4, SbMlo8, SbMlo9, SbMlo11*, *SbMlo13*), CAT element in two genes (*SbMlo6*, *SbMlo7*), CATT in only *SbMlo*, CTAG in *SbMlo11*, CG-motif in *SbMlo12*, GATA in *SbMlo9*, GA-motif in *SbMlo11* and *SbMlo13*, GT-1 motif in *SbMlo6* and *SbMlo12*, GAG-motif in *SbMlo8* and *SbMlo13*, GCC in *SbMlo2*, GARE and GTGGC motifs in *SbMlo6*, I-box in four genes (*SbMlo5, SbMlo9, SbMlo12, SbMlo13*), MBS in three genes (*SbMlo4, SbMlo5*, *SbMlo12*), O2-site in three genes (*SbMlo6, SbMlo12, SbMlo13*), Sp1 in eight genes (*SbMlo1, SbMlo2, SbMlo5, SbMlo6, SbMlo8, SbMlo9 SbMlo12¸SbMlo13*), Skn-1 in *SbMlo4-9* and *SbMlo13*, LTR in three genes (*SbMlo2, SbMlo4, SbMlo11*), MNF-1 in *SbMlo4,* TGACG in five genes (*SbMlo1, SbMlo4-6, SbMlo8, SbMlo10, SbMlo13*), TATC in *SbMlo11,* TCCACCT-motif in four genes (*SbMlo2, SbMlo3, SbMlo11*, *SbMlo13*), TCA-element in three genes (*SbMlo1, SbMlo5, SbMlo12*), TATCCAT/C in *SbMlo2,* TGA element in *SbMlo4,* TCT in *SbMlo5,* Motif IIb in three genes (*SbMlo4*, *SbMlo6, SbMlo12*), plant AP2-like in *SbMlo4,* and circadian related element in three of the *Mlo* genes (*SbMlo6, SbMlo7,SbMlo9*).

Thus G-Box was the most frequent *cis*-elements found in 10 of the *Mlo* genes; Sp1, was the next most abundant element occurring in eight genes, while CGTCA, CCGTCC were present in six genes each. Other elements were found in the upstream sequences of five or less number of the *Mlo* genes.

Fig. 5- Cis-acting elements study of 12 *SbMlo* genes using PLANTCARE

Since all the 15 *A. thaliana Mlo* genes are fully characterized, we retrieved -1000 upstream regions of these genes to find out their putative *cis*-acting elements and compared them with the *cis*-elements identified in the upstream regions of the 12 putative *Mlo* genes of *S. bicolor*. The predicted *cis*-acting elements for *AtMlo*s and their expression data retrieved using TAIR database (http://www.arabidopsis.org/) are summarized in Table 5. The *SbMlo*s showing sequence homology to the *AtMlo*s are also listed.

Primers for PCR Amplification in *Sorghum*

From complete genomic sequences of the 13 identified SbMlo genes, 24 primer pairs were designed to amplify full length genes from *S. bicolor* (Supplementary table 1). Further, to study the expression of the SbMlo genes, 13 primer pairs were designed using primer3 [31] on the basis of the predicted RNA encoded by the SbMlo gene CDS (Supplementary table 2).

Supplementary Table 1- Genomic primer pairs for amplification of identified full length genes.

Primer name	Primer sequence	Ln	Tm	GC%	COMPL	Product Size
SbMlo1F1	CCACTTTGATCCTCCCCTTT	20	60.3	50	2	2207bp
SbMlo1R1	CTGCACGCGAGAAGATGTAA	20	60.16	50		
SbMlo1F2	AGGGCGTCTGAAAGTAAGCA	20	60.02	50	0	3205bp
SbMlo1R2	TGCACGTTTGTTCAGAGAGG	20	60.02	50		
SbMlo2F1	CAATCGTCTCATCTGCTCCA	20	59.94	50	3	2022bp
SbMlo2R1	CCGAACAGATCGGAAGAACT	20	59.28	50		
SbMlo2F2	GACCCAAAACGTCTCCAGAA	20	60.09	50	1	3011bp
SbMlo2R2	CCGATTTCACTCGGCTTTAC	20	59.71	50		
SbMlo3F1	CTAATCGGTTCCGCTCTCAA	20	60.34	50	2	2051bp
SbMlo3R1	TTCTGTGCAACCTCAACAGC	20	60.03	50		
SbMlo3F2	GCTGTTGAGGTTGCACAGAA	20	60.03	50	0	2012bp
SbMlo3R2	GGAATCCCCACTCCATCATA	20	59.56	50		
SbMlo4F1	ACGTATATTGGGCCAACGAC	20	59.71	50	2	2165bp
SbMlo4R1	TGGAATCTCTGCCAATACCC	20	59.89	50		
SbMlo4F2	GGAAACACTGGGAGGATGAA	20	59.9	50	3	2540bp
SbMlo4R2	CCTAAACACATCGCCTCCAT	20	59.96	50		
SbMlo5F1	GGAGGACAAGATCCAGCAAA	20	60.2	50	2	2503bp
SbMlo5R1	CCCATGGCAGACATGTTGTA	20	60.39	50		
SbMlo6F1	GCCTCGATCCGCATTAACTA	20	60.2	50	1	2313bp
SbMlo6R1	CCTTGTTATGGAGCCCTTGA	20	60.07	50		
SbMlo6F2	TCTAGCTGCGGAACCAATTC	20	60.35	50	0	3008bp
SbMlo6R2	GGCCTTAGCAAATCATCAGC	20	59.81	50		
SbMlo7F1	CAACGGTCCATGACCATACA	20	60.24	50	1	2047bp
SbMlo7R1	TAGAGCGAAACACACGATGC	20	60.02	50		
SbMlo7F2	CATAGATCGGCGGTATGCTT	20	60.08	50	1	2506bp
SbMlo7R2	GAGCTCGGCCTCTCAAATTA	20	59.55	50		
SbMlo8F1	GATCCCCCATTTTCCTTCTC	20	59.7	50	1	2578bp
SbMlo8R1	ACGTACAAGCCACATGACCA	20	60.03	50		
SbMlo8F2	TGAATCTTGCAGGGTGACTG	20	59.83	50	2	2714bp
SbMlo8R2	GAGAAGCCATTTCTCGTCCA	20	60.34	50		
SbMlo9F1	ATACCCGATGGGGATTTAGC	20	60.01	50	0	1855bp
SbMlo9R1	CACCTTACCAGCCAACACAA	20	59.61	50		
SbMlo9F2	GACCCAGCTTGTTGTTGTTG	20	59.19	50	0	2525bp
SbMlo9R2	CTCCACTCATGCATTCCTTG	20	59.24	50		
SbMlo10F	TGCTTTTAGGAGGTGCACAG	20	59.07	50	2	3160bp
SbMlo10R	AAGGCTAACGGGCCTAAACT	20	59.29	50		
SbMlo11F1	TCCTTCCAACCAAACACTCC	20	59.94	50	1	3123bp
SbMlo11R1	AGCTCCTTAAGGCTCCGTTAGT	22	59.94	50		
SbMlo11F2	TGTACAACGCTGGACCATTC	20	59.57	50	0	3176bp
SbMlo11R2	TGCTCTCGGCAAAGACCTAT	20	59.98	50		
SbMlo12F1	GGGAGTGTATGCTTGCCTTT	20	59.2	50	1	2998bp
SbMlo12R1	GACCCAAATTCCTGGAGATG	20	59.34	50		
SbMlo12F2	CCATGCAAGAGAGACAACCA	20	59.83	50	0	2836bp
SbMlo12R2	CAGGAAGCAACTCGTCACAA	20	60.02	50		
SbMlo12F3	GTGCCTTTTCACTGGAAACC	20	59.57	50	0	2795bp
SbMlo12R3	CTTTTTGTCCCCTTCCTCCT	20	59.55	50		
SbMlo13F	GCCCAGAAATCGGAGAGAAT	20	60.55	50	0	3210bp
SbMlo13R	CTGCATACTGTCTGCTAAGGTACG	24	60.38	50		

Supplementary Table 2- Primer for expression studies of SbMlo genes.

Primer name	Primer sequence	Tm	GC	Base pairs	Product Size
SbMlo 1F	GTGGAGCCAAGTGACAGGTT	60.16	55	20	202
SbMlo 1R	AGAGGACCTGAAGAGCCACA	59.99	55	20	
SbMlo 2F	AGGCATGCCAGTGGTAAAAC	60	50	20	298
SbMlo 2R	TGTCTGCTCGTCGAAAATTG	59.99	45	20	
SbMlo 3F	AGGCATTGGGAAGGAAAGAT	59.9	45	20	221
SbMlo 3R	CCCCTGCAGTGTTTCTCAAT	60.11	50	20	
SbMlo 4F	CTCCCTCATCGTGGTCATCT	60.07	55	20	267
SbMlo 4R	ATTCCTTGGAACACCGTCAG	59.97	50	20	
SbMlo 5F	AGTCGTTGGCATAAGCTGGT	59.76	50	20	168
SbMlo 5R	GCAACCTCAAGAGCCAACTC	60	55	20	
SbMlo 6F	TGGCTCACAACTCAAACTGC	60.03	50	20	152
SbMlo 6R	TGCTGTAGCTGCACCAAAAC	60.06	50	20	
SbMlo 7F	TTAGTTCCGTTGGCTCTGCT	60.02	50	20	212
SbMlo 7R	AACTCGAAGGCGTTCTGAAA	59.99	45	20	
SbMlo 8F	CGTGAGGGGTACGAGTCATT	59.99	55	20	283
SbMlo 8R	AAACAAGCCACATGACCACA	60.01	45	20	
SbMlo 9F	TTGGCACACGTACTTTTGGA	60.15	45	20	226
SbMlo 9R	AAACGCGTTCTGAAAGAGGA	59.99	45	20	
SbMlo 10F	CTCATGCATCATGGGACAAG	60.07	50	20	243
SbMlo 10R	ATCGTCGATTCGTTCTTGCT	59.84	45	20	
SbMlo 11F	CACACAGCTATCGAGGACGA	60.01	55	20	284
SbMlo 11R	TAGCAGCTCCCCATCTGAGT	59.97	55	20	
SbMlo 12F	GCAAAGGGATCTTGGATGAA	60.01	45	20	247
SbMlo 12R	TCAGCAACAGGAAGACGATG	59.98	50	20	
SbMlo 13F	CAACCTTCGTGCTCTTCCTC	59.99	55	20	122
SbMlo 13R	CGAGTCCATCCCGTACTCAT	59.95	55	20	

Discussion

Characteristics of *S. bicolor Mlo* genes

The 13 predicted *SbMlo* genes are distributed on 8 chromosomes as provided in the chromosome data of. *S. bicolor* genome in NCBI datamodel. *Sorghum* chromosome 9 has 3 *SbMlo* genes, chromosomes 10 and chromosome 1 have 2 genes each, while chromosomes 2, 3, 4, and 5 have 1 gene each. The complete catalogue of Mlo proteins in a single plant species is useful for viewing the existing sequential, structural and functional diversity associated with its diverse roles played in plants. The evolutionary relationships between different Mlo proteins were analyzed by subjecting the deduced amino acid sequences encoded by the identified 13 *SbMlo* genes for multiple sequence alignment. Amino acid composition of these hypothetical Mlo proteins showed that they all are leucine-rich. It also provides evidence that *Mlo* gene family belongs to leucine-rich class of plant disease resistance genes. Multiple sequence alignment of these SbMlo proteins showed that they are a well conserved family divided into two subgroups containing 3 clusters. The analysis of introns/exon gene structures revealed that most introns have conserved positions and phases, providing the evidence for the intron–early theory, and that multiple independent intron loss events are likely to have occurred during evolution of flowering plants. The hypothesis that genome wide and tandem duplication contributed to the expansion of the *Mlo* gene family across the plant kingdom seems to be applicable for *Sorghum* as well as two other diploid species, namely, *Arabidopsis* and *Oryza*, which contain 15 and 12 *Mlo* genes, respectively [21].

Evolutionary relationship among *S. bicolor*, *O. sativa* and *A. thaliana Mlo* genes

Evolutionary relationship of the 13 *Sorghum Mlo* genes with the 12 *O. sativa* [22], and 15 *A. thaliana* genes (obtained from complete genome of the species) revealed a similar classification pattern in sequence evolution in these three species. The 15 members of the *Arabidopsis Mlo* gene family are well characterized, and they have been shown to mainly function as modulators of plant defence and cell death. The coding regions of the 13 putative *SbMlo* genes show substantial coding sequence homology with one or the other member of the *Arabidopsis Mlo* gene family; therefore, the *SbMlo* genes are also expected to play an important role in the defense mechanism of *Sorghum* [3,8,23,36] *Mlo* genes are novel Calmodulin-binding Proteins [20]. Some members of *Arabidopsis Mlo* gene family are also reported to play an important role during leaf senescence, seedling development, stigma receptivity, pollen tube development, fruit ripening, and development of flower bud and flower abscission zone [7].

Evolutionary study of *SbMlo genes* revealed two groups (A and B) of which group A could be divided into three clades. Clade III contained *SbMlo1, 2* and *13*, which exhibited similarity with *AtMlo9*, *10*, *5, 7* and *8*. Clade II contains *SbMlo5 and 3*, which showed homology with *AtMlo12, 2, 6* and *3*. Clade I consisted of *SbMlo 7, 4, 9, 10, 11* and *12,* which were grouped with *AtMlo1, 13 and 15*. Clade IV contained *SbMlo6* and *8* that showed similarity with *AtMlo4*, *11* and *14*. Cluster I of Group A consisted of *SbMlo1*, 13 and 2, which has been classified with *AtMlo2*, *AtMlo6*, *AtMlo12*, *AtMlo3*, *AtMlo9*, *AtMlo10*, *AtMlo5, AtMlo7* and *AtMlo8*. Promoter region comparison between *Arabidopsis* and *Sorghum Mlo* genes showed possible role of the *SbMlos* during defense response to fungus, seedling development stage and leaf senescence with modulatory function as reported in the case of *Arabidopsis* genes [7,9,18]. Cluster -II group A had *SbMlo10*, *SbMlo11*, Sb*Mlo*9, Sb*Mlo*4,Sb*Mlo*5, Sb*Mlo*7, Sb*Mlo*3 and Sb*Mlo*12, which have been classified with At*Mlo*1, At*Mlo*15 and At*Mlo*13; thus these *SbMlo* genes may play a role during early seedling growth, cotyledon vascular system development, in pollen and in papillae [7,9,37]. Studies on Group B cluster-III revealed that Sb*Mlo*8 and Sb*Mlo*6 were closely similar with At*Mlo*4, At*Mlo*11 and At*Mlo*14. Sb*Mlo*8 and Sb*Mlo*6 may play an important role during seedling growth, flower development, fruit abscission as per phylogenetic classification and promoter comparison with *Arabidopsis*.

Insilico functional study from *Sorghum Mlo* genes

The *cis*-regulatory element analysis of the predicted *SbMlo* genes revealed major putative function as regulation of genes associated with abiotic and biotic stresses, photoperiod response, growth hormone and meristem specific elements. The occurrence of CAAT, TATA and G-Boxes is very high in the upstream regions of the predicted *SbMlo.* Genes *SbMlo5, 10*, *4*, *5, 6* and *12* showed MYB binding site involved in flavonoid biosynthetic gene regulation. Skn-1 element required for endosperm expression was observed in *SbMlo4* to *SbMlo9*. TTGAC (WBOXATNPR1), TCA (Salicylic acid responsive), TGAC (WBOXNTERF3), TCT (light responsive), TATCCAT/C, TGACG (methyl jasmonate), TCCACCT, CCGTCC elements were also observed in the upstream regions, which may play a role in regulation of genes involved in defense mechanism of *Sorghum* plants.

Table 5- Study of cis-acting elements in upstream region of AtMlo genes and their function based on GUS activity patterns. The elements common in SbMlo and AtMlo promoter region are highlighted in green color.

Gene (*A. thaliana*)	Expression during (as shown by GUS activity patterns)	Cis-acting elements (Promoter/ Silencer/Enhancer) in -1000 upstream sequences	Coding region (exonic) sequence similarity with *S. bicolor*
AtMlo1	Early seedling growth, in root and cotyledon vascular system, in pollen and in papillae.	5UTR Py-rich stretch, AE-box, ARE, ATCT-motif, Box4, CAAT-box, GA-motif, GARE-motif, GCN4_motif, I-box, MBS, P-box, Skn-1_motif, TATA-box, TC-rich repeats, TCA –element, TCT-motif, chs-CMA1a	*SbMlo4,SbMlo9, SbMlo10, SbMlo11*
AtMlo 3	Early seedling growth, in primary root and lateral root primordia, in fruit abscission zone, in vascular system of cotyledons and in trichomes of young leaves. It was not expressed in mature rosette leaves	A-box, AAGAA-motif, AE-box, ARE, Box4, BoxI, CAAT-box, CCGTCC-box, CGTCA-motif, F-box, G-Box, HD-Zip1, HD-Zip2, HSE,LTR, Skn-1_motif, TATA-box, TC-rich repeats, TCA-element, TCT-motif, TGACG-motif, chs-CMA2b, circadian	*SbMlo1*
AtMlo 4	Early seedling growth, in roots and lateral root primordia, in flower and fruit abscission zone, in vascular system of root, cotyledons and young leaves. It was not expressed in mature rosette leaves	5UTR Py-rich stretch, AAGAA-motif, ABRE, ACE, ARE, Box-W1, CAAT-box, CATT-motif, CCAAT-box, G-Box, GT1-motif, Gap-box, MBS, MRE, MSA-like, Sp1, TATA-box, TC-rich repeats, TCA-element, TGA-element, W box	*SbMlo6**
AtMlo 5	Seedling growth, in cotyledon vascular system, and in stigma, anther and pollen grains. It was not expressed in rosette leaves	AAGAA-motif, AC-II, ACE, ARE, CAAT-box, CCAAT-box, CTAG-motif, G-Box, GA-motif, GAG-motif, GARE-motif, HSE, LAMP-element, TATA-box, TATC-box, TC-rich repeats, TCA-element, TCCC-motif, TCT-motif, TGA-element	*SbMlo2*
AtMlo 6	Early seedling growth, in roots and lateral root primordia, in flower and fruit abscission zone, in vascular system of cotyledons, young leaves and petals, in mature rosette leaves and in anthers	AE-box, ARE, ATCT-motif, BoxI, CAAT-box, CGTCA-motif, ERE, GA-motif, GARE-motif, GATA-motif, GT1-motif, LTR, Nodule-site2, Skn-1_motif, TATA-box, TCA-element, TGACG-motif	*SbMlo1*
AtMlo 7	In vegetative organs (RT-PCR experiments)and in pollen grains	ABRE, ACE, AE-box, ARE, AT-rich element, ATCT-motif, Box4, BoxI, CAAT-box, ELI-box3, G-box, HSE, MRE, SKn-1_motif, TATA-box, TC-rich repeats	*SbMlo3*
AtMlo 8	Seedling growth, in cotyledons and hypocotyl, and in fruit abscission zone	3-AF1 binding site, 5UTR Py-rich stretch, AAGAA-motif, AE-box, ARE, ATGCAAAT motif, Box4, CAAT-box, GAG-motif, GATA-motif, HSE, Skn-1_motif, HSE, Skn-1_motif, TATA-box, TC-rich repeats, TCA-element, TCT-motif	*SbMlo13*
AtMlo 9	Early seedling growth, in cotyledon vascular system, in flowers (with strong expression in anthers) in siliques and fruit abscission zone; not expressed in roots, or in mature rosette leaves	5UTR Py-rich stretch, Box4, BoxI, CAAT-box, CGTCA-motif, ERE, O2-site, TA-rich region, TATA-box, TC-rich repeats, TCT-motif, TGA-element, TGACG-motif, WUN-motif, circadian	*SbMlo2*
AtMlo 10	in root and cotyledon vascular system, in root-shoot junction and lateral root primordia and in developing siliques	3-AF1 binding site, 3-AF3 binding site, 5UTR Py-rich stretch, AAGAA-motif, ABRE, AE-box, AT-rich element, CAAT-box, CCAAT-box, CGTCA-motif, G-box, GAG-motif, GATA-motif, Gap-box, I-box, MNF1, MRE, P-box, Skn-1_motif, TATA-box, TATC-box, TC-rich repeats, TCA-element, TCT-motif, TGACG-motif, circadian	*SbMlo2*
AtMlo 11	During early seedling growth, in root tips and cotyledon vascular system, in floral organs (anthers and stigma), and in fruit abscission zone	AAGAA-motif, AE-box, ARE, BoxI, C-repeats/DRE, CAAT-box, CGTCA-motif, ERE, GA-motif, GAG-motif, GARE-motif, GCN4_motif, LTR, MBS, O2 -site, Skn-1_motif, TATA-box, TGA-box, TGACG-motif	*SbMlo8**
AtMlo 12	During early seedling growth, in root tips and cotyledon vascular system, in floral organs (anthers and stigma), and in fruit abscission zone	3-AF1 binding site, AAGAA-motif, ABRE, AE-box, ARE, CAAT-box, CAT-box, CATT-motif, CGTCA-motif, G-box, HSE, I-box, LTR, MBS, O2-site, Skn-1_motif, TATA-box, TC-rich repeats, TCCACCT-motif, TGA-element, TGACG-motif, circadian, rbcS-CMA7a	*SbMlo1*
AtMlo 13	Early seedling growth, in root and cotyledon vascular system, in pollen and also in placenta of developing siliques	5UTR Py-rich stretch, AAGAA-motif, ARE, AT-rich element, Box4, CAAT-box, CGTCA-motif, E2Fa, G-box, GA-motif, I-box, LAMP-element, LTR, MBS, Skn-1_motif, TATA-box, TC-rich repeats, TCA-element, TGACG-motif, box S, circadian, sbp-CMA1c	*SbMlo12**
AtMlo 14	Early seedling growth, in developing primary root, and particularly in root tips of 10-day old seedlings; it was not expressed in leaves or flowers	5UTR Py-rich stretch, AAGAA-motif, AE-box, ARE, Box4, CAAT-box, CCAAT-box, CGTCA-motif, GARE-motif, GCN4_motif, GT1-motif, LTR, MBS, Skn-1_motif, TATA-box, TC-rich repeats, TGACG-motif, circadian	*SbMlo8*
AtMlo 15	Early seedling growth, in root tips and flower (papillae, anthers and pollen grains)	AAGAA-motif, ACE, ARE, Box4, BoxI, Box-W1, CAAT-box, CATT-motif, ERE, G-Box, GATA-motif, GT1-motif, Gap-box, MBS, MRE, Skn1_motif, TATA-box, TC-rich repeats, TCA-element, TCCC-motif, W box, box S, circadian	*SbMlo4, SbMlo9, SbMlo10, SbMlo11*

Table 6- Cis-acting Elements, their sequences and function

S.N.	Cis-acting Element	Sequence	Function
1	AAGAA-motif	gGTAAAGAAA	unknown
2	ABRE	TACGTG	cis-acting element involved in the abscisic acid responsiveness
3	Box 4	ATTAAT	part of a conserved DNA module involved in light responsiveness
4	Box III	CATTTACACT	protein binding site
5	CAAT-box	CAAT	common Cis-acting element in promoter and enhancer regions
6	G-Box	CACGTG	cis-acting regulatory element involved in light responsiveness
7	GAG-motif	GGAGATG	part of a light responsive element
8	GC-motif	CCCCCG	enhancer-like element involved in anoxic specific inducibility
9	GCC box	AGCCGCC	unknown
10	I-box	GATAAGATA	part of a light responsive element
11	Skn-1_motif	GTCAT	cis-acting regulatory element required for endosperm expression
12	Sp1	CC(G/A)CCC	light responsive element
13	TATA-box	TAATA	core promoter element around -30 of transcription start
14	TC-rich repeats	ATTCTCTAAC	cis-acting element involved in defense and stress responsiveness
15	Box-W1	TTGACC	fungal elicitor responsive element
16	MBS	TAACTG	MYB binding site involved in drought-inducibility
17	MNF1	GTGCCC(A/T)(A/T)	light responsive element
18	TGACG-motif	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
19	plant_AP-2-like	CGCGCCGG	Unknown
20	GATA-motif	AAGGATAAGG	part of a light responsive element
21	HSE	AAAAAATTTC	cis-acting element involved in heat stress responsiveness
22	LTR	CCGAAA	cis-acting element involved in low-temperature responsiveness
23	MSA-like	TGCAACGGC	cis-acting element involved in cell cycle regulation
24	TCA-element	TCAGAAGAGG	cis-acting element involved in salicylic acid responsiveness
25	TGA-element	AACGAC	auxin-responsive element
26	ACE	AAAACGTTTA	cis-acting element involved in light responsiveness
27	LAMP-element	CCTTATCCA	part of a light responsive element
28	rbcS-CMA7a	GGCGATAAGG	part of a light responsive element
29	TATCCAT/C-motif	TATCCAT	Unknown
30	box S	AGCCACC	Unknown
31	AE-box	AGAAACAT	part of a module for light response
32	F-box	CTATTCTCATT	Unknown
33	WUN-motif	TCATTACGAA	wound-responsive element
34	A-box	CCGTCC	cis-acting regulatory element related to meristem specific activation
35	AT1-motif	AATTATTTTTTATT	part of a light responsive module
36	circadian	CAANNNNATC	cis-acting regulatory element involved in circadian control
37	GARE-motif	TCTGTTG	gibberellin-responsive element
38	GC-motif	CCCCCG	enhancer-like element involved in anoxic specific inducibility
39	HD-Zip 3	GTAAT(G/C)ATTAC	protein binding site
40	3-AF1 binding site	TAAGAGAGGAA	light responsive element
41	O2-site	GATGACATGA	cis-acting regulatory element involved in zein metabolism regulation
42	P-box	CCTTTTG	gibberellin-responsive element
43	ATCT-motif	AATCTAATCC	part of a conserved DNA module involved in light responsiveness

Conclusion

In this study, a comprehensive computational analysis was conducted, and 13 members of the *Mlo* gene family were identified in *Sorghum*. A complete overview of this gene family in *Sorghum* is presented, including the multiple sequence alignment, gene structures, phylogeny, chromosomal locations and their *cis*-regulatory element analysis. The comparative phylogenetic analysis with respect to the *Mlo* gene family clearly indicated the proximity of *Sorghum Mlos* with rice and *Arabidopsis Mlo* genes, even when *Sorghum* and rice are monocots, while *Arabidopsis* is a dicot. Further, the presence of similar groups and subgroups in comparative phylogeny of *Sorghum*, rice and *Arabidopsis Mlo* genes indicates conservation of *Mlo* gene sequences even in widely separated taxonomic groups of plants. The identified proteins showed similarity with signature accession PF03094 (Pfam database), IPR004326 (INTERPROSCAN), cellular component integral to membrane (GO: 0016021) and biological process with cell death (GO: 0008219) from Gene Ontology database (http://www.geneontology.org/). The *in silico* investigation of putative genes from *Sb Mlo* gene family needs to be supported by wetlab experiments through expression profiling of respective genes by designing the molecular markers using hypothetical mRNA and amplification of full length candidate *Mlo* genes from *S. bicolor*.

Acknowledgement

The facilities provided by DBT-funded SUB-DIC, Centre for Bioinformatics, Faculty of Science, Banaras Hindu University, Varanasi are thankfully acknowledged.

References

[1] Altschul S.F., Gish W., Webb Miller Eugene W. Myers and Lipman D.J. (1990) *Journal of Molecular Biology*, 215,403-410.

[2] Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W. and Lipman D.J. (1997) *Nucleic Acids Research*, 25, 3389-3402.

[3] Ascencio-Ibáñez J.T., Sozzani R., Lee T.J., Chu T.M., Wolfinger R.D., Cella R., Hanley-Bowdoin L. (2008) *Plant Physiol*. 148, 436-54.

[4] Bailey T.L. and Gribskov M. (1998) *Bioinformatics*, 14, 48-54.

[5] Bailey T.L., Williams N., Misleh C. and Li W.W. (2006) *Nucleic Acids Research*, 34, W369-W373.

[6] Büschges R., Hollricher K., Panstruga R., Simons G., Wolter M., Frijters A., Van Daelen R., Van der Lee T., Diergaarde P., Groenendijk J., Töpsch S., Vos P., Salamini F. and Schulze-Lefert P. (1997) *Cell*. 88, 695-705.

[7] Chen Z., Hartmann H.A., Wu M.J., Friedman E.J., Chen J.G., Pulley M., Schulze-Lefert P., Panstruga R., Jones A.M. (2006) *Plant Mol. Biol*., 60, 583-97.

[8] Chen Z., Noir S., Kwaaitaal M., Hartmann H.A., Wu M.J., Mudgil Y., Sukumar P., Muday G., Panstruga R., Jones A.M. (2009) *Plant Cell*, 21, 1972-91.

[9] Consonni C., Bednarek P., Humphry M., Francocci F., Ferrari S., Harzen A., Ver Loren van Themaat E., Panstruga R. (2010) *Plant Physiol*. 152, 1544-61.

[10] Consonni C., Humphry M.E., Hartmann H.A., Livaja M., Durner J., Westphal L., Vogel J., Lipka V., Kemmerling B., Schulze-Lefert P., Somerville S.C., Panstruga R. (2006) *Nat. Genet*., 38, 716-20.

[11] Devoto A., Hartmann H.A., Piffanelli P., Elliott C., Simmons C., Taramino G., Goh C.S., Cohen F.E., Emerson B.C., Schulze-Lefert P., Panstruga R. (2003) *J. Mol. Evol*., 56(1), 77-88.

[12] Devoto A., Piffanelli P., Nilsson I., Wallin E., Panstruga R., von Heijne G., Schulze-Lefert P. (1999) *J. Biol. Chem.,* 274, 34993-5004.

[13] Elliott C., Zhou F., Spielmeyer W., Panstruga R., Schulze-Lefert P. (2002) *Mol. Plant Microbe. Interact*. 15(10), 1069-77.

[14] Falquet L., Pagni M., Bucher P., Hulo N., Sigrist C.J., Hofmann K., Bairoch A. (2002) *Nucleic Acids Research,* 30, 235-238.

[15] FAO (2010) *http://www.fao.org/corp/statistics/en/.*

[16] Feechan A., Jermakow A.M., Torregrosa L., Panstruga R. and Dry I.B. (2008) *Functional Plant Biology,* 35, 1255-1266.

[17] Finn R.D., Mistry J., Schuster-Bockler B., Griffiths-Jones S., Hollich V., lassmann T., Moxon S., Marshall M., Khanna A., Durbin R., Eddy S.R., Sonnhammer E.L.L. and Bateman A. (2006) *Nucleic Acids Research*, 34, D247-D251.

[18] Humphry M., Bednarek P., Kemmerling B., Koh S., Stein M., Göbel U., Stüber K., Pislewska-Bednarek M., Loraine A., Schulze-Lefert P., Somerville S., Panstruga R. (2010) *Proc. Natl. Acad. Sci*. 107, 21896-901.

[19] Jorgensen J.H. (1992) *Euphytica,* 63, 141-152.

[20] Kim M.C., Panstruga R., Elliot C., Muller J., Devoto A., Yoon H.W., Park H.C., Cho M.J., Schulze-Lefert P. (2002) *Nature* 416, 447-450.

[21] Konishi S., Sasakuma T. and Sasanuma T. (2010) *Genes Genet. Syst.,* 85, 167-175.

[22] Liu Q. and Zhu H. (2008) *Gene*, 409, 1-10.

[23] Ma S. and Bohnert H.J. (2007) *Genome Biol*., 8, R49.

[24] Marta S. Lopes, Jose Luis Araus, Philippus D.R. Van Heerden and Christine H. Foyer (2011) *J. Exp. Bot.,* 62, 3135-3153.

[25] Panstruga R. (2005) *Plant Mol. Biol*., 59(3), 485-500.

[26] Piffanelli P., Zhou F., Casais C., Orme J., Jarosch B., Schaffrath U., Collins N.C., Panstruga R., Schulze-Lefert P. (2002) *Plant Physiol.,* 129(3), 1076-85.

[27] Pruitt K.D., Tatusova T. and Maglott D.R. (2007) *Nucleic Acids Research*, 35, D61-D65.

[28] Quevillon E., Silventoinen V., Pillai S., Harte N., Mulder N., Apweiler R. and Lopez R. (2005) *Nucleic Acids Research*, 33, W116-W120.

[29] Schauser L., Wieloch W., Stougaard J. (2005) *J. Mol. Evol.,* 60, 229-237.

[30] Solovyev V., Kosarev P., Seledsov I., Vorobyev D. (2006) *Genome Biol.*, 7(S10.) 1-12.

[31] Steve Rozen and Helen J. Skaletsky (2000) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*., 365-386.

[32] Tamura K., Peterson D., Peterson N., Stecher G., Nei M. and Kumar S. (2011) *Molecular Biology and Evolution,* 28, 2731-2739.

[33] Thompson J.D., Higgins D.G., Gibson T.J. (1994) *Nucleic Acids Res*., 22, 4673-4680.

[34] Tusnády G.E. and Simon I. (2001) *Bioinformatics,* 17(9), 849-50.

[35] Varshney R.K., Hoisington D.A., Tyagi A.K. (2006) *Trends Biotechnol*. 24, 490-499.

[36] Vogel J., Somerville S. (2000) *Proc. Natl. Acad. Sci*., 97, 1897-902.

[37] Wang Y., Zhang W.Z., Song L.F., Zou J.J., Su Z., Wu W.H. (2008) *Plant Physiol.,* 148, 1201-11.

[38] Wolter M., Hollricher K., Salamini F., Schulze-Lefert P. (1993) *Mol. Gen. Genet*., 239, 122-128.