Genome-wide characterization of MLO family genes in bread wheat and Arabidopsis thaliana shed light on their role in growth and development, and resistance to powdery mildew, biotic and abiotic stresses

Babar Hussain[1], Hamza Ramzan[2,3], Qasim Raza[4,5], Rana Muhammad Atif [6,7*], Hikmet Budak[8], Abdel-Halim I. Ghazy[9]

[1]Deaprtment of Biotechnology, Faculty of Science and Technology, University of Central Punjab, Lahore, Pakistan
[2] Max Planck Institute for Plant Breeding Research, Cologne, Germany.
[3] Hungarian University of Agriculture and Life Sciences, Godollo, Hungary.
[4] College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China.
[5]Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Weifang, Shandong 261325, China.
[6]Precision Agriculture and Analytics Lab, Centre for Advanced Studies in Agriculture and Food Security, National Centre in Big Data and Cloud Computing, University of Agriculture Faisalabad, Faisalabad, Pakistan
[7]Department of Plant Breeding and Genetics, University of Agriculture Faisalabad, Faisalabad, Pakistan
[8] Western Arizona College, Yuma, Arizona, United States
[9]Plant Production Department, College of Food and Agriculture Sciences, King Saud University, P.O. Box 2460, Riyadh 11451, Saudi Arabia

*Corresponding author e-mail: Rana Muhammad Atif dratif@uaf.edu.pk

Abstract

Powdery mildew (PM), caused by the fungus, Blumeria graminis tritici is a devastating and notorious disease that is prevalent across the globe and causes up to 62% of yield losses in wheat. The well-known 68 PM resistance loci (Pm1 - Pm68) are not long-lasting and break down when new races of pathogens interact with host plants. The editing of three known MLO genes through CRISPR/Cas9 has conferred resistance against PM. However, only seven MLO genes were known in wheat to date. Taking advantage of IWGSC Ref-seq v2.1 and using the gold-standard analysis pipeline that we have designed, we expanded this number to 47 and reported 40 additional potential CRISPR targets that can be used to develop PM resistance, disease resistance, and abiotic stress tolerance in wheat. We also performed phylogenetic, gene structure, and MOTIF analysis that indicated conserved evolutionary patterns, and the phylogenetic pattern was strongly supported by gene structure and MOTIF analysis. The gene duplication and evolution analysis showed that segmental duplications were prevalent and purifying selection operated during evolution. Comparative phylogenetic analysis

with Arabidopsis thaliana and Brachypodium distachyon depicted that clade IV was exclusive for monocots. Finally, MLOs are expressed under powdery mildew, rusts, head blight, abiotic stresses, growth, and development thus TraesMLOs have a potential role in wheat under these conditions. Therefore, we hope that the MLO genes identified in this study will be edited through CRISPR/Cas9 or will be overexpressed to develop PM and disease-resistant and abiotic stress-tolerant wheat. Additionally, we performed most of the above-mentioned analyses for Arabidopsis thaliana, revised its numbers as compared to the TAIR database, and updated it outdated analysis performed in 2003.

Key words: Wheat, *Arabidopsis thaliana*, *Brachypodium distachyon*, MLO, Powdery mildew, evolution, phylogenetic analysis, abiotic stresses

1. Introduction

Climate change not only alters temperature and rain patterns but it results in the emergence of new races of pathogens. Several fungal pathogens lower wheat yield by causing diseases such as leaf rust, stripe rust, stem rust, smuts, septoria leaf blotch, fusarium head blight, and powdery mildew (Hussain et al., 2022). Powdery mildew (PM), caused by the fungus, *Blumeria graminis*, is a devastating and notorious disease that is prevalent across the globe and causes up to 62% of yield losses in wheat (Singh et al., 2016). Therefore, efforts have been made to identify resistance genes against this pathogen. A total of 68 resistance genes (officially designated *Pm1* - *Pm68*) were mapped in wheat after decades-long efforts (McIntosh et al.; Hussain et al., 2022). However, resistance against ever-evolving pathogens is never long-lasting and breaks down when new races of pathogens interact with host plants (Scholthof, 2007; Hussain, 2015).

Keeping in view the economic impact of this disease, ten genes were cloned in wheat to develop PM resistance till date i.e., (*Pm1a, Pm2, Pm3, Pm3b, Pm5e, Pm21, Pm24, Pm38/Lr34/Yr18/Sr57, Pm41, Pm46/Lr67/Yr46/Sr55, Pm60*) (Yahiaoui et al., 2004; Krattinger et al., 2009; Hurni et al., 2013; Moore et al., 2015; Sánchez-Martín et al., 2016; Xing et al., 2018; Zou et al., 2018; Li et al., 2020; Lu et al., 2020; Xie et al., 2020). However, cloning techniques such as map-based cloning are very difficult to perform, are tedious, time-consuming, and need specialized lab facilities and skilled manpower, therefore, plants' race to adapt against evolving pathogens is unsuccessful in terms of slow progress in developing PM resistance varieties.

However, the onset of the CRISPR revolution in 2012 (Jinek et al., 2012) and its subsequent use for the editing of powdery mildew susceptibility genes, *Mildew Locus O* (*MLO*), conferred resistance against PM as reported by two independent groups (Shan et al., 2014; Wang et al., 2014). TALEN-based editing of three homologues of a MLO gene and CRISPR/Cas9-based editing of *TaMLO-A1* allele resulted in increased resistance to PM (Wang et al., 2014). Recently,

CRISPR/Cas9 technology was used to introduce a 304-kilobase pair targeted deletion in *TaMLO-B1* locus that resulted in robust PM resistance without any growth penalties. This was consequence of altered chromatin that led to the ectopic activation of *Tonoplast monosaccharide transporter 3* (*TaTMT3B*), and negated the negative effects of MLO disruption (Li et al., 2022). As CRISPR can be used to edit the effector-binding elements of susceptibility genes that directly interact with pathogens (Hussain et al., 2018; Hussain and Ahmad, 2022), it provides durable and broad-spectrum resistance against diseases. However, only seven MLO genes were known till now (Konishi et al., 2010), but, there are 15 MLO genes in *Arabidopsis thaliana* (Devoto et al., 2003), 15 in diploid lentil (Polanco et al., 2018), 39 in soybean (Deshmukh et al., 2014), and 68 in octaploid strawberry (Tapia et al., 2021), so we hypotheses that being a hexaploid, bread wheat should have at least 40-45 MLO genes.

In this study, we performed a genome-wide search to identify all MLO genes in bread wheat, studied their distribution in the genome, performed phylogenetic, gene structure, MOTIF, gene duplication, and evolution analysis to establish their locations in genome and evolutionary relationship; and finally, their expression under powdery mildew, other diseases, abiotic stresses, and growth and development to predict their potential roles in wheat under these conditions. We hope that additional MLO genes identified in this study will be edited through CRISPR/Cas9 to improve PM and resistance against other disease, and growth and development in addition to developing abiotic stress tolerance in wheat.

Additionally, using the latest reference genome i.e., TAIR 10, we corrected several mistakes in gene numbers of 15 MLO genes in *Arabidopsis thaliana* (Devoto et al., 2003) and compared it with wheat and *Brachypodium distachylon*. We also expanded the number of genes for *Brachypodium distachylon* and used them for a comparative phylogenetic tree construction as old analysis became outdated after the non-availability of MLO sequences.

2. Methodology

2.1. Sequence Retrieval and Genome-wide identification of MLO Genes

The *T. aeastivum* genome assemblies, proteomes, transcriptomes along with general feature format files (GFF3) were downloaded from the International Wheat Genome Sequencing Consortium v2.1 (IWGSC RefSeq v2.1) (Zhu et al., 2021) website (https://urgi.versailles.inrae.fr/download/iwgsc/IWGSC_RefSeq_Assemblies/v2.1/). A total of 15

Arabidopsis thaliana MLO sequences were downloaded from The Arabidopsis Information Resource (TAIR) and the genome, proteome, CDS and annotation files were downloaded from TAIR as well. The genome of *Brachypoidum distachyon* was obtained from Ensemble Plant database. Genome-wide identification of MLO gene family was performed for *Triticum aeastivum. Arabidopsis thaliana* MLO sequences were used as query sequences for the genome-wide search of MLO genes within the genomes of *Triticum aeastivum (Traes)* through BLAST. Additionally, an HMMER-based search was performed to predict the MLO proteins in wheat genomes (Potter et al., 2018). The presence of complete MLO superfamily domain in the obtained proteins was confirmed through the NCBI Conserved Domain Database (CDD) search tool v3.19 (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and Pfam/Interpro (http://pfam.xfam.org/). Th proteins without complete MLO conserved domain were not considered for further analysis as previously described (Raza et al., 2022). Moreover, Expasy server's pI/Mw tool (https://web.expasy.org/compute_pi/) was used for the determination of biophysical characteristics like molecular weight and isoelectric point of the MLO peptides. The subcellular localization of proteins was predicted with a web-based server Bologna Unified Subcellular Component Annotator (BUSCA) (https://busca.biocomp.unibo.it/). For the purpose, the organ that proclaimed the highest value was taken as the location of the protein.

2.2. Chromosomal Mapping/Distribution across the genome

The identified genes were mapped on the chromosomes by using TB Tools software (Chen et al., 2020). The chromosomal location of genes extracted from GFF3 file.

2.3. Phylogenetic Analysis

The identified MLO sequences were aligned through multiple sequence alignment (MSA) tool i.e., Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) (Sievers et al., 2011) with default settings. The resultant multiple sequence alignment was used to construct a phylogenetic tree with the maximum likelihood (ML) method using MEGA v11.0 (Kumar et al., 2018).

2.4. Comparative Phylogenetic Analysis

The identified TraesMLO sequences along with 15 *AtMLO* sequences, and 12 BradiMLO proteins were aligned through Clustal Omega and the rest of procedure was the same as described above.

2.5. Gene Structure and Motif Analysis

The coding sequences (CDS) along with respective genomic sequences for all *TraesMLO* genes were used in the Gene Structure Display Server (GSDS) tool (http://gsds.gao-lab.org/) to visualize the exons, introns, and down and upstream regions of the genes. Conserved motifs/domain analysis was performed by online tool MEME 5.1 (http://meme-suite.org/tools/meme). We tried different settings for the proper identification of conserved motifs and found a suitable setting of 16 as maximum number of motifs and 50 as maximum length of motifs. The motif data was downloaded as an XML file to be used by TB tools software (Chen et al., 2020) to visualize and synchronize the gene structure.

2.6. Divergence (Gene Duplication) and Evolution Analysis

We followed the methodology described by Raza and colleagues (Raza et al., 2022) for the purpose. The Coding sequences (CDS) of all *TraesMLO* genes were subjected to MSA using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) (Sievers et al., 2011) and aligned sequences were MAFFT aligned (pairwise) in all possible combinations using Sequence Demarcation Tool v1.2 (Muhire et al., 2014) to calculate the sequence identities. Gene pairs with 90% identity (E value < $1e^{-10}$) were classified as duplicated. The duplicated gene pairs that were mapped on the same chromosome were considered as tandem duplication and the ones located on different chromosomes were defined as segmental duplication. The CDS of duplicated genes were MAFFT aligned before being subjected to the Synonymous Non-Synonymous Analysis Program v2.1.1 (www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html) to compute the synonymous (Ks) and non-synonymous (Ka) substitutions. The Ka/Ks was also calculated to predict the codon selection during evolution. The calculation for an approximate divergence time for the duplicated genes was done by using formulae $T = Ks/2r \times 10{-}6$ assuming a substitution rate (r) of $6.5 \times 10{-}9$ substitutions/synonymous site/year (El Baidouri et al., 2017)**.**

2.7. Expression Analysis

RNA-seq based expression patterns of all identified MLO genes during the whole growth and development period, as well as under different abiotic (cold, drought, heat and phosphate starvation) and biotic (fusarium head blight, powdery mildew and stripe rust) stresses were retrieved from the expVIP Wheat Expression Browser as Log2 TPM (processed expression value in transcripts per million (Borrill et al., 2016; Ramírez-González et al., 2018). Expression heatmaps were generated with TBtools (Chen et al., 2020).

3. Results and Discussion

3.1. A Gold-Standard Pipeline for Discovery of MLOs

3.1.1. Bread Wheat

Using the bread wheat Refseq genome v2.1, we initially identified 61 genes encoding MLO superfamily proteins using MLO protein sequences from *Arabidopsis thaliana* as queries by usig BLAST function of TB tools. The main sequential steps of the pipeline are: a). Two duplicate genes were deleted by sorting the genes on the basis of their names. b). Multiple sequence alignment identified four pseudogenes with incomplete domain and such proteins are reported to lose their function (Panchy et al., 2016). c). phylogenetic tree helped to identify 8 isomers thus leaving behind 47 non-redundant MLO genes. d). NCBI CDD search and e). Pfam/Interpro (http://pfam.xfam.org/) was used to confirm the presence of MLO super family principal domain in the protein sequence of these genes.

Following the confirmation, renaming of these TraesMLO genes was done in ascending order according to their positions on the chromosomes (TraesMLO1 to TraesMLO47) following the renaming in other species (Deshmukh et al., 2014; Kusch et al., 2016; Win et al., 2018). The peptide length of TraesMLO proteins ranged from 409 to 612 amino acids and average length was 514 amino

acids (**Table 1**). The isoelectric point values varied from 7.36 to 9.87, while the molecular weight of the proteins was in range of 46.98 to 68.50 kDa, with a mean kDa value of 58.34 (**Table 1**).

Several other groups reported a varying number of MLO superfamily members in different plant species. For example, 15 MLO genes were identified in Arabidopsis (Devoto et al., 2003), whereas 20 in pumpkin (Win et al., 2018), 15 in lentil (Polanco et al., 2018), 68 in octaploid strawberry (Tapia et al., 2021), 39 in soybean (Deshmukh et al., 2014) etc. As several species have 15 MLO genes therefore, wheat has 47 MLO genes being a hexaploid.

3.1.2. *Arabidopsis thaliana*

Using the *Arabidopsis thaliana* latest reference genome version, TAIR 10 as a subject sequence and a TaesMLO protein as a query, we were able to identify 33 MLO proteins. However, after subjecting the sequences of these proteins to our pipeline (3.1.1), we were able to narrow them down to 15 MLOs that are also available on TAIR database. However, use of latest version of the genome meant that we were able to assign new names of AtMLOs available on TAIR database. We renamed these genes as per their chromosomal locations as a standard procedure described for

several other species (Deshmukh et al., 2014; Kusch et al., 2016; Win et al., 2018). The new names, gene start, end, sizes, and protein sizes and other properties of MLOs are given in **Table 2**.

3.1.3. *Brachypodium distachyon*

While trying to get sequences of MLOs from Bradi Genome, we found out that analysis was performed on a very old version of genome and protein sequences were not available (Ablazov and Tombuloglu, 2016). Therefore, by using a AtMLO protein sequence as a query, we were able to identify and expand the MLOs from 11 to 12 for *Brachypodium distachyon* and used these proteins for comparative phylogenetic tree. We renamed these genes as per their chromosomal locations as a standard procedure described for several other species (Deshmukh et al., 2014; Kusch et al., 2016; Win et al., 2018). The new names, gene start, end, sizes, and protein sizes and other properties of MLOs are given in **Table 3**.

Table 1. Detailed properties of identified MLO genes in *Triticum aestivum* (AABBDD genome).

S#	Gene Name/ID	New Gene Names	LINKAGE Group	Gene Start	Gene End	Gene Length (bp)	Strand	pI	Mw (kDa)	Protei Lengt (AA)
1	TraesCS1A03G0539600.1	*TraesMLO1*	Chr1A	364505363	364509878	4515	-	9.36	63.32	56
2	TraesCS1A03G0668200.1	*TraesMLO2*	Chr1A	457568166	457572624	4458	-	9.08	56.52	49
3	TraesCS2A03G0163000.1	*TraesMLO3*	Chr2A	42897580	42900950	3370	-	9.45	55.86	49
4	TraesCS2A03G0163100.1	*TraesMLO4*	Chr2A	42905631	42908828	3197	-	9.63	55.33	48
5	TraesCS2A03G0775400.1	*TraesMLO5*	Chr2A	539693546	539696928	3382	-	9.03	55.33	49
6	TraesCS2A03G1299000.1	*TraesMLO6*	Chr2A	768633437	768638338	4901	-	9.22	68.50	60
7	TraesCS3A03G0013800.1	*TraesMLO7*	Chr3A	6965742	6973314	7572	-	8.51	61.25	54
8	TraesCS3A03G0944700.1	*TraesMLO8*	Chr3A	648479170	648482130	2960	-	9.33	54.81	48
9	TraesCS4A03G0580300.1	*TraesMLO9*	Chr4A	518674297	518679167	4870	+	8.62	58.39	50
10	TraesCS5A03G0963900.1	*TraesMLO10*	Chr5A	599592378	599596257	3879	+	9.12	62.37	56
11	TraesCS5A03G1150400.1	*TraesMLO11*	Chr5A	661126976	661129967	2991	-	9.87	61.05	54
12	TraesCS5A03G1158600.1	*TraesMLO12*	Chr5A	664743006	664756436	**13430**	+	7.36	64.06	58
13	TraesCS5A03G1159400.1	*TraesMLO13*	Chr5A	664797864	664801136	3272	-	9.53	60.29	53
14	TraesCS6A03G0387200.1	*TraesMLO14*	Chr6A	154203217	154207538	4321	+	8.70	56.92	49
15	TraesCS6A03G0805500.1	*TraesMLO15*	Chr6A	545643614	545648358	4744	-	9.28	54.72	48
16	TraesCS7A03G0928000.1	*TraesMLO16*	Chr7A	561794230	561801544	7314	-	9.51	54.31	47
17	TraesCS1B03G0628500.1	*TraesMLO17*	Chr1B	398073082	398077619	4537	-	9.41	63.28	56
18	TraesCS1B03G0759600.2	*TraesMLO18*	Chr1B	484447774	484452526	4752	-	9.15	56.25	49
19	TraesCS2B03G0227100.1	*TraesMLO19*	Chr2B	65216529	65219835	3306	-	9.52	55.85	49
20	TraesCS2B03G0227200.1	*TraesMLO20*	Chr2B	65315357	65318657	3300	-	9.59	55.36	48
21	TraesCS2B03G0856800.1	*TraesMLO21*	Chr2B	479483631	479486919	3288	+	8.99	55.06	49
22	TraesCS2B03G1489500.1	*TraesMLO22*	Chr2B	788988904	788994031	5127	-	9.38	68.44	61
23	TraesCS3B03G1076900.1	*TraesMLO23*	Chr3B	689913560	689916559	2999	+	9.34	55.05	48
24	TraesCS4B03G0222000.1	*TraesMLO24*	Chr4B	104220694	104226387	5693	+	7.97	59.19	51
25	TraesCS4B03G0835400.1	*TraesMLO25*	Chr4B	611711110	611728680	**17570**	+	9.72	60.87	54
26	TraesCS4B03G0836900.1	*TraesMLO26*	Chr4B	612012622	612014655	2033	-	9.49	60.39	53
27	TraesCS5B03G1014800.1	*TraesMLO27*	Chr5B	590155469	590160172	4703	+	9.00	59.26	53
28	TraesCS6B03G0494800.1	*TraesMLO28*	Chr6B	232292016	232296365	4349	-	8.55	56.56	49
29	TraesCS6B03G0955000.1	*TraesMLO29*	Chr6B	601764897	601769827	4930	-	9.14	56.54	49
30	TraesCS7B03G0775700.1	*TraesMLO30*	Chr7B	524484720	524491473	6753	-	9.14	57.36	50
31	TraesCS1D03G0519600.1	*TraesMLO31*	Chr1D	293427154	293431832	4678	-	9.42	63.60	56
32	TraesCS1D03G0630400.1	*TraesMLO32*	Chr1D	357744682	357749428	4746	-	9.09	56.56	49
33	TraesCS2D03G0160400.1	*TraesMLO33*	Chr2D	37249636	37253058	3422	-	9.53	55.81	49
34	TraesCS2D03G0160500.1	*TraesMLO34*	Chr2D	37272260	37275524	3264	-	9.76	54.30	49
35	TraesCS2D03G0715400.1	*TraesMLO35*	Chr2D	400923755	400926943	3188	-	8.60	55.34	49
36	TraesCS2D03G1262400.1	*TraesMLO36*	Chr2D	638659281	638665452	6171	-	9.28	68.46	61
37	TraesCS3D03G0010400.1	*TraesMLO37*	Chr3D	1805922	1810130	4208	-	9.06	50.84	44
38	TraesCS3D03G0877800.1	*TraesMLO38*	Chr3D	513351507	513355165	3658	-	9.41	55.19	48
39	TraesCS4D03G0182500.1	*TraesMLO39*	Chr4D	69109133	69114152	5019	+	8.93	57.68	50
40	TraesCS4D03G0743900.1	*TraesMLO40*	Chr4D	483173859	483176970	3111	+	9.77	60.55	54
41	TraesCS4D03G0744500.1	*TraesMLO41*	Chr4D	483205801	483208560	2759	-	9.52	60.45	53
42	TraesCS5D03G0918700.1	*TraesMLO42*	Chr5D	481711884	481717428	5544	+	9.07	60.50	54

43	TraesCS6D03G0340700.1	*TraesMLO43*	Chr6D	148637851	148642268	4417	-	8.80	56.91	49
44	TraesCS6D03G0681600.1	*TraesMLO44*	Chr6D	418198229	418205340	7111	-	9.35	46.98	40
45	TraesCS7D03G0041500.1	*TraesMLO45*	Chr7D	8641641	8683479	**41838**	-	9.83	53.97	46
46	TraesCS7D03G0062000.1	*TraesMLO46*	Chr7D	14007248	14014362	7114	+	9.69	54.81	48
47	TraesCS7D03G0891400.1	*TraesMLO47*	Chr7D	493179947	493187259	7312	-	9.15	57.52	50

Table 2. Detailed properties of identified MLO genes in model plant, *Arabidopsis thaliana*, genome

S#	Gene/Protein IDs	TAIR IDs	Chr.	New IDs	Gene Start	Gene End	Gene Size (bp)	pI	MW (kDa)	Protein length (aa)
1	AT1G11000.1	*AtMLO4*	1	*AtMLO1*	3671756	3676133	4377	8.18	65.47	573
2	AT1G11310.1	*AtMLO2*	1	*AtMLO2*	3800500	3804233	3733	9.87	65.54	573
3	AT1G26700.1	*AtMLO14*	1	*AtMLO3*	9227378	9232437	5059	8.56	63.92	554
4	AT1G42560.1	*AtMLO9*	1	*AtMLO4*	15980135	15983571	3436	9.45	53.21	460
5	AT1G61560.1	*AtMLO6*	1	*AtMLO5*	22708468	22712289	3821	10.06	67.23	583
6	AT2G17430.1	*AtMLO7*	2	*AtMLO6*	7568018	7570912	2894	9.67	62.11	542
7	AT2G17480.1	*AtMLO8*	2	*AtMLO7*	7590295	7594195	3900	9.52	67.21	593
8	AT2G33670.1	*AtMLO5*	2	*AtMLO8*	14244951	14248155	3204	9.46	56.90	501
9	AT2G39200.1	*AtMLO12*	2	*AtMLO9*	16356041	16359943	3902	9.27	66.55	576
10	AT2G44110.2	*AtMLO15*	2	*AtMLO10*	18245357	18248459	3102	9.4	56.32	497
11	AT3G45290.1	*AtMLO3*	3	*AtMLO11*	16617398	16621370	3972	7.56	58.56	508
12	AT4G02600.1	*AtMLO1*	4	*AtMLO12*	1143888	1147562	3674	8.69	59.13	526
13	AT4G24250.1	*AtMLO13*	4	*AtMLO13*	12574719	12577583	2864	9.76	55.07	478
14	AT5G53760.1	*AtMLO11*	5	*AtMLO14*	21822323	21826488	4165	9.21	66.35	573
15	AT5G65970.1	*AtMLO10*	5	*AtMLO15*	26387428	26390767	3339	9.38	64.48	569

Table 3. Detailed properties of identified MLO genes in *Brachypodium distachyon* genome

S#	Gene Ids	Chr	New Gene Names	Gene Start	Gene End	Gene size (bp)	Protein IDs	pI	MW (kDa)	Prot. Size (AA)
1	BRADI_1g36172v3	1	*BradiMLO1*	31956564	31964215	7651	XP_003560590.1	9.21	57.72	511
2	BRADI_1g76500v3	1	*BradiMLO2*	73314560	73319326	4766	XP_003558907.1	9.54	63.01	556
3	BRADI_2g25190v3	2	*BradiMLO3*	23087065	23092426	5361	XP_003568437.1	9.17	56.13	493
4	BRADI_2g33517v3	2	*BradiMLO4*	33413404	33418187	4783	XP_003568790.1	9.01	62.42	572
5	BRADI_2g57317v3	2	*BradiMLO5*	55280734	55285025	4291	XP_003564808.1	9.15	56.26	500

6	BRADI_3g07150v3	3	*BradiMLO6*	5310788	5315931	5143	XP_003571073.1	8.45	36.34	496
7	BRADI_3g32230v3	3	*BradiMLO7*	34235029	34239882	4853	XP_003574266.1	9.33	63.72	561
8	BRADI_3g46070v3	3	*BradiMLO8*	47756268	47759741	3473	XP_003575162.1	9.82	56.04	495
9	BRADI_4g08741v3	4	*BradiMLO9*	7875919	7880542	4623	XP_024319424.1	9.18	50.16	452
10	BRADI_4g23887v3	4	*BradiMLO10*	28654097	28661227	7130	XP_003577771.1	8.74	59.28	513
11	BRADI_5g11205v3	5	*BradiMLO11*	14917808	14922517	4709	XP_003581276.1	9.00	56.20	501
12	BRADI_5g26435v3	5	*BradiMLO12*	27620517	27625925	5408	XP_024312147.1	8.97	67.52	600

3.2. Distribution of MLOs in bread wheat genome/s

Total 16, 14, and 17 MLO genes were mapped on A, B, and D Genomes of Traes, respectively in non-random fashion (**Figure 1A, B, C**). Such non-random chromosomal distribution of genes has been reported in other species as well (Ahmed et al., 2021). Nonrandom distribution of plant gene families originates from long terminal repeat retrotransposons through evolution (Gao et al., 2004). Chromosomes 2A, 2B, and 2D harbored maximum four TraesMLO genes each while chromosome 1A, 1B, 1D, 6A, 6B, and 6D contained two genes each. Interestingly, chromosome 5A harbored four TraesMLO genes while 5B and 5D contained only one gene. Similarly, chromosome 4A harbored only one gene in contrast to three TraesMLO genes on both 4B and 4D. Additionally, the number of mapped genes on 7D were three in comparison to one gene on 7A and 7B while two, one, and two genes were mapped on chromosome 3 in three genomes, respectively (**Figure 1A, B, C**).

Previously, seven MLO genes were identified in wheat and none of these genes showed change in their expression under abiotic stresses (Konishi et al., 2010). This could be linked with non-availability of reliable wheat reference genome till that date. However, the availability of whole genome refence sequence of bread (Appels et al., 2018; Zhu et al., 2021) has provided us with an opportunity to scan whole genome and were able to identify 47 MLO genes. Additionally, these seven MLO gene members were not assigned to chromosomes due to non-availability of annotated genome (Konishi et al., 2010) that we have shown here.

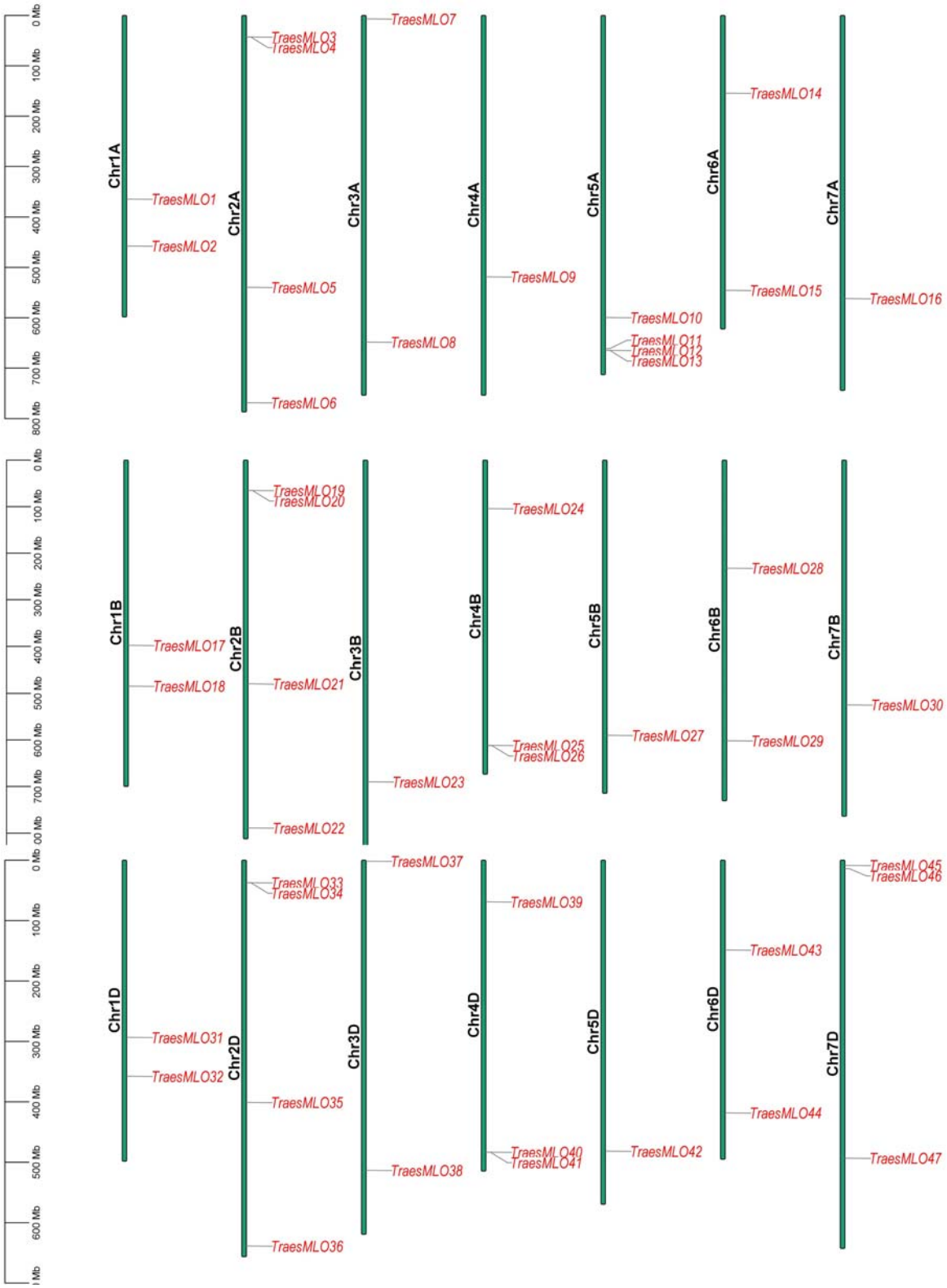

Figure 1: Chromosomal mapping of 47 bread wheat MLO genes on all 21 chromosomes of wheat

3.3. Phylogenetic Classification of wheat and *Arabidopsis thaliana* MLO family

We constructed a maximum likelihood phylogenetic tree to demonstrate the evolutionary relationship of the MLO members in modern wheat. The phylogenetic tree showed the distribution of MLO proteins into three distinct groups (**Fig. 2a**). The distribution of MLO superfamily members into three groups is because group-specific motifs are conserved among the proteins (**Figure 2a, c**). Group I represent the largest group with 28 members out of 47 MLO proteins identified in bread wheat, but physiological roles of most of these proteins are be to be investigated. Group two and three were relatively small. Total thirteen TraesMLO proteins were present in Group II while Group III had only six TraesMLO proteins in it (**Fig. 2a**). The maximum likelihood phylogenetic analysis of *Arabidopsis thaliana* MLO proteins strongly supported the phylogenetic group classification of wheat i.e., Group I had 9 proteins as it is a diploid species so it had three times a smaller number of genes in it (**Figure 3a**). While group II and III had total three genes each and it each group has specific conserved MOTIFs in it (**Figure 3a, c**). Additionally, wheat and Arabidopsis phylogenetic trees were further divided into 7 distinct subgroups/clades (**Figure 2a; 3a**). This classification pattern of MLO proteins into three main groups and seven subgroups/clades is well documented in soybeans (Deshmukh et al., 2014), pumpkin (Win et al., 2018), lentil (Polanco et al., 2018) and strawberry (Tapia et al., 2021).

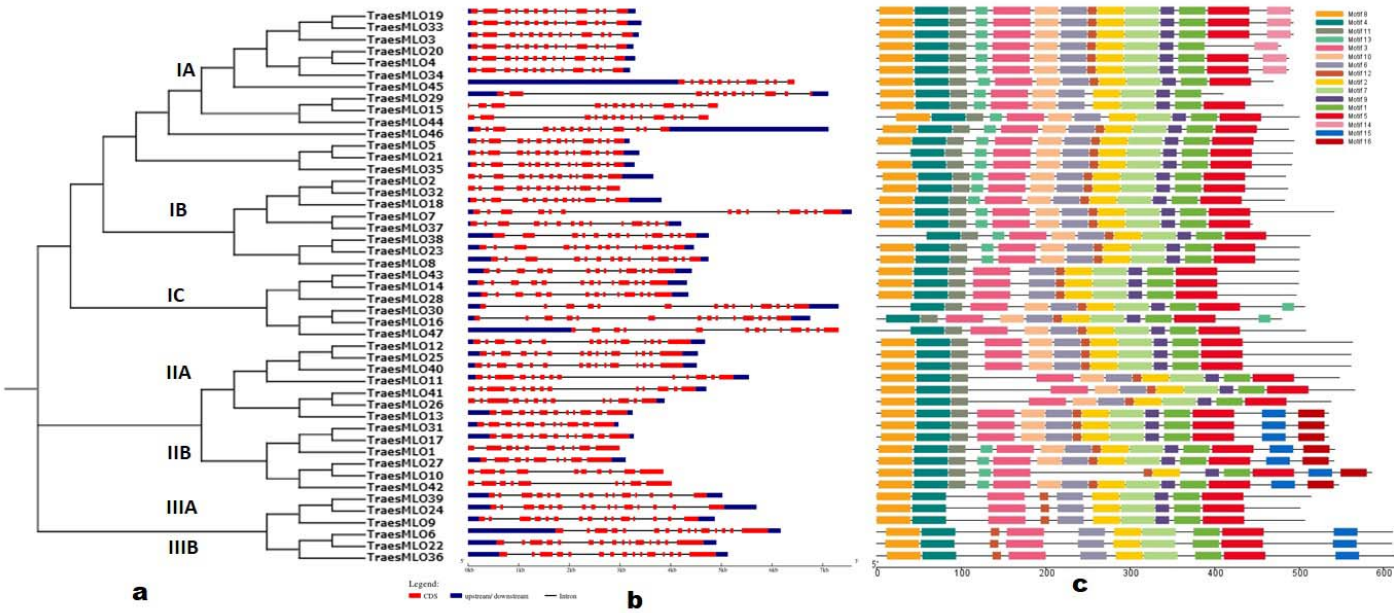

Figure 2: Comparison of maximum likelihood phylogenetic tree of 47 TraesMLO genes with gene structure and conserved motifs of TraesMLO genes/proteins

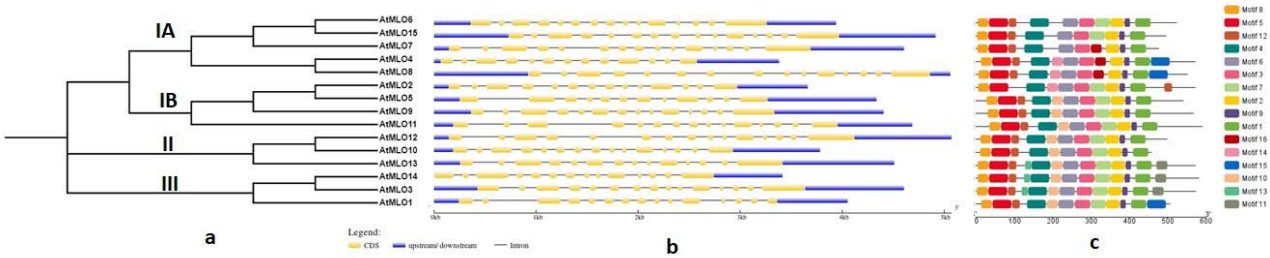

Figure 3: Comparison of maximum likelihood phylogenetic tree of 15 AtMLO proteins with gene structure and conserved motifs of AtMLO genes/proteins

3.4. Gene Structure and Conserved Domain/MOTIFs Analysis

3.4.1. Comparison of Gene Structure and MOTIFs with Phylogenetic Tree

Different exon and intron structures are of great importance for plant biodiversity and variation among gene families. These distinct structures play an important role in the fate of genes (Xu et al., 2012). Our Gene Structure analysis revealed that *TraesMLO* genes have introns ranging from 9 to 15 (**Fig 2b**). Similar exon/intron patterns have been outlined in bitter gourd (*Momordica charantia*) with 8 to 17 introns (Chen et al., 2021) and bottle gourd (*Lagenaria siceraria*) with 5 to 15 introns (Wang et al., 2023). On the other hand, the gene structure analysis in melon (*Cucumis melo*) showed the range of introns from 1 to 16 (Zhang et al., 2023). Total 21 Out of 47 *TraesMLO* genes shared the same 13 number of introns. Furthermore, results revealed that only one gene TraesMLO40 contains 9 introns, *TraesMLO11*, *TraesMLO12* and *TraesMLO25* contains 10 introns, *TraesMLO13*, *TraesMLO26*, *TraesMLO41* and *TraesMLO45* contains 11 introns, *TraesMLO44* contains 12 introns, 8 genes contain 14 introns, and 9 genes contains 15 introns (**Fig 2b**).

It was also noticed that 6 genes (*TraesMLO11*, *TraesMLO12*, *TraesMLO15*, *TraesMLO23*, *TraesMLO25*, *TraesMLO29*) have no upstream and downstream regions, 3 genes (*TraesMLO10*, *TraesMLO27* and *TraesMLO38*) without upstream and 2 genes (*TraesMLO16* and *TraesMLO45*) without downstream region. In comparison with the phylogenetic tree, we found that all the members of group III in phylogenetic tree have same intron/exon distribution patterns that confirmed the evolutionary phylogenetics in our studies. All the members of phylogenetic

subgroups IIA and IIIB also follow the same pattern of intron/ exons patterns (**Fig 2b**). Secondly, after motif analysis we found that group I also shows differentiation within its subgroups, namely IA, IIB, and IC based on conserved motifs (**Fig 2c**). Furthermore, some motifs clearly differentiate the groups and subgroups based on their presence and absence. For example, motif 11 with 21 amino acids (ESAMSHMLPCKPPPHSVAAEG) was absent in group III but present in group II (**Fig 2a, b, c**).

Similarly, if we see the distinctness within the group then motif 16 with 32 amino acids (AHPVHRVPPCDGWRSASSPALDAHIPGADFCF) was differentiating between the subgroups IIA and IIB as it is absent in subgroup IIA but present in all the member of subgroup IIB (**Fig 2a, b, c**). Overall, our gene structure analysis and the patterns of motifs strongly support evolutionary phylogenetic relationships in newly identified MLO proteins.

3.4.2. Sequences and Characteristics of Identified MOTIFs

The MEME online server was used to perform conserved domain analysis to further figure on the diversity in the structure of TraesMLO genes. A total number of 16 conserved domains were identified among 47 TraesMLO genes (**Table 4; Fig 2c**). The width of the conserved domains was found to be ranging from 11 (Motif 12) to 50 (Motif 5) amino acids (**Table 4**), while the different motifs are illustrated with different colors in Figure 3. Within the groups, the nature, number, and patterns of motifs were observed almost same but variations among the groups and subgroups can easily be seen. For example, in group III, there are 9-10 motifs that are highly conserved, while in group I, there are 11-14 motifs that are conserved. Furthermore, group I also shows differentiation within its subgroups, namely IA, IIB, and IC based on conserved motifs (**Fig 2c**). Furthermore, some motifs clearly differentiate the groups and subgroups based on their presence and absence. For example, motif 11 with 21 amino acids (ESAMSHMLPCKPPPHSVAAEG) was absent in group III but present in group II (**Table 4**; **Fig 2c**).

Similarly, if we see the distinctness within the group then motif 16 with 32 amino acids (AHPVHRVPPCDGWRSASSPALDAHIPGADFCF) was differentiating between the subgroups IIA and IIB as it is absent in subgroup IIA but present in all the member of subgroup IIB (**Table 4**; **Fig 2c**). The four motifs, Motif 1 with 32 amino acids (LIHFILFQNAFEIAFFFWILVTYGFKSCIMGN), motif 2 with 33 amino acids (YMIRALEDDFKKVVGISWYLWAFVVIFLLLNVN), motif 3 with 45 amino acids

(VPLLSLEAJHQLHIFIFVLAVTHVLFSAITMLLGGAKIRQWKKWE) and motif 4 with 41 amino acids (KRHKKALYEALLKVKEELMLLGFISLLLTVTQGPISKICIP) are conserved in all the 47 TraesMLO genes that demonstrate the basic structural and functional resemblances among all the genes (**Table 4**; **Fig 2c**). It is reported that MLO gene belongs to a group of plant disease susceptible genes with leucine rich characteristics (Singh et al., 2012). We found that all the putative 47 TraesMLO genes are leucine rich. Furthermore, it was also observed that each member of the TraesMLO genes adheres to the seven-transmembrane structure, a significant hallmark of the MLO gene family (Devoto et al., 2003). While not all MLO genes across various species follow this specific structure, they are still classified as MLO genes. For example, in soybean, among the 39 identified MLO genes, only 18 possess the precise configuration of seven transmembrane domains; others depict variations with more or fewer transmembrane domains (Deshmukh et al., 2014).

Table 4: Multi-level consensus sequences for the MEME defined domains observed among different TaMLO proteins.

Motif	Multilevel Consensus Sequence	E-value	No. of Sites	*Width
1	LIHFILFQNAFEIAFFFWILVTYGFKSCIMGN	7.1e-1067	47	32
2	YMIRALEDDFKKVVGISWYLWAFVVIFLLLNVN	3.5e-1016	47	33
3	VPLLSLEAJHQLHIFIFVLAVTHVLFSAITMLLGGAKIRQWKKWE	3.0e-1360	47	45
4	KRHKKALYEALLKVKEELMLLGFISLLLTVTQGPISKICIP	4.4e-1192	47	41
5	YVIARLVJGVIVQVLCSYSTLPLYAJVTQMGSSFKKSIFDEQVQEGLVNW	4.7e-1362	45	50
6	ILSWMRSFFKQFYGSVTKEDYTTLRLGFIMKH	5.0e-905	46	32
7	GWHTYFWISFVPLILLLVVGTKLEHIITELAYEVAZKHAAI	2.2e-1070	46	41
8	GGGEARTLELTPTWIVASVCSVIVIISLLFERLLHRLGKWL	4.8e-949	42	41
9	GDLVVKPSDDHFWFGRP	7.9e-427	44	17
10	ENAGNGPSMIRHVHZFKFIRDHFKGHGKR	2.4e-416	37	29
11	ESAMSHMLPCKPPPHSVAAEG	2.10E-233	41	21
12	LRGNPKFBFHK	4.10E-153	44	11
13	ARRLLAGGAASDDHC	7.60E-127	27	15
14	EPIDEANGGAVQMTNTRANSSVEQGTARL	2.10E-102	6	29
15	QPSRGSSPVHLLHKAGARSDDPQSAPPSP	1.60E-102	10	29
16	AHPVHRVPPCDGWRSASSPALDAHIPGADFCF	4.60E-101	7	32

3.5. Comparative Phylogeny of wheat MLOs with *Arabidospsi thaliana* and *Brachypodium Distachyon*

Comparative phylogenetic analysis was performed by using the 15 reported proteins of *Arabidpsis thaliana*, 12 proteins of *Brachypodium distachyon* (expanded from previous reported 11) and newly identified 47 MLO proteins of *Triticum asestivum* to deep dive into the evolutionary patterns and functional confirmation of MLO proteins. As a result, we found that all the proteins are distributed into seven clades (**Fig 4**). Similarly, seven clade pattern has been reported in soybeans (Deshmukh et al., 2014), pumpkin (Win et al., 2018), lentil (Polanco et al., 2018) and strawberry (Tapia et al., 2021). We classified the clades according to the pattern of Arabidopsis reported in Devoto et al., 2003. In the Clade I, *AtMLO14, AtMLO3* were distributed with 2 *BradiMLO* and 6 *TraesMLO* genes; and Clade II contained *AtMLO13, AtMLO12, and AtMLO10* genes with 4 BradiMLO and 14 *TraesMLO* genes, Clade III contained *AtMLO1, AtMLO8, AtMLO7, AtMLO6, AtMLO15* genes with 3 BradiMLO and 6 TraesMLO genes, and Clade V contained *AtMLO2, AtMLO8, AtMLO9* with BradiMLO02 and 7 TraesMLO genes (**Figure 4**). Interestingly, there was not a single gene that clustered in the clade with *AtMLO11* but we kept this as clade VI keeping in mind the previous studies that supported seven clades pattern in monocots especially as reported in cereal model species- *Brachypodium distachyon* (Ablazov and Tombuloglu, 2016).

It is possible that the absence of genes exhibiting a similar phylogeny to AtMLO11 in wheat and *Brachypodium distachyon* might be attributed to difference between Arabidopsis and wheat with respect to monocot and dicot point of view. Furthermore, the intriguing identification of Clade IV exclusively containing 13 TraesMLO genes with 02 BradiMLO genes, devoid of any AtMLO genes, provides a valuable insight into the monocot-specific nature of this clade as reported previously as well (Win et al., 2018). This observation may potentially indicate distinct functional roles and evolutionary paths that are specific to monocotyledonous species, separate from their dicotyledonous counterparts. In addition, we used *Brachypodium distachyon*, a model species in cereals, solely for the purpose of verifying the absence of AtMLO genes within this clade. It is thinkable that this clade could be specific to cereals, thus offering valuable insights to researchers in discerning the evolutionary history and phylogenetics of cereals. In conclusion, our comparative phylogenetic analysis of MLO proteins in *Arabidopsis thaliana, Brachypodium distachyon* and *Triticum aestivum* has yielded valuable insights into the evolutionary patterns and functional implications of these proteins. The consistent seven-clade distribution pattern across diverse plant species suggests a deep-rooted and robust evolutionary framework. The alignment with well-

established Arabidopsis patterns, as well as the observed monocot-specific clades, adds further layers of complexity and specificity to the understanding of MLO protein evolution. This study contributes to the broader field of plant molecular evolution and lays the foundation for future investigations into the functional significance of these identified clades in the context of disease resistance and plant development.

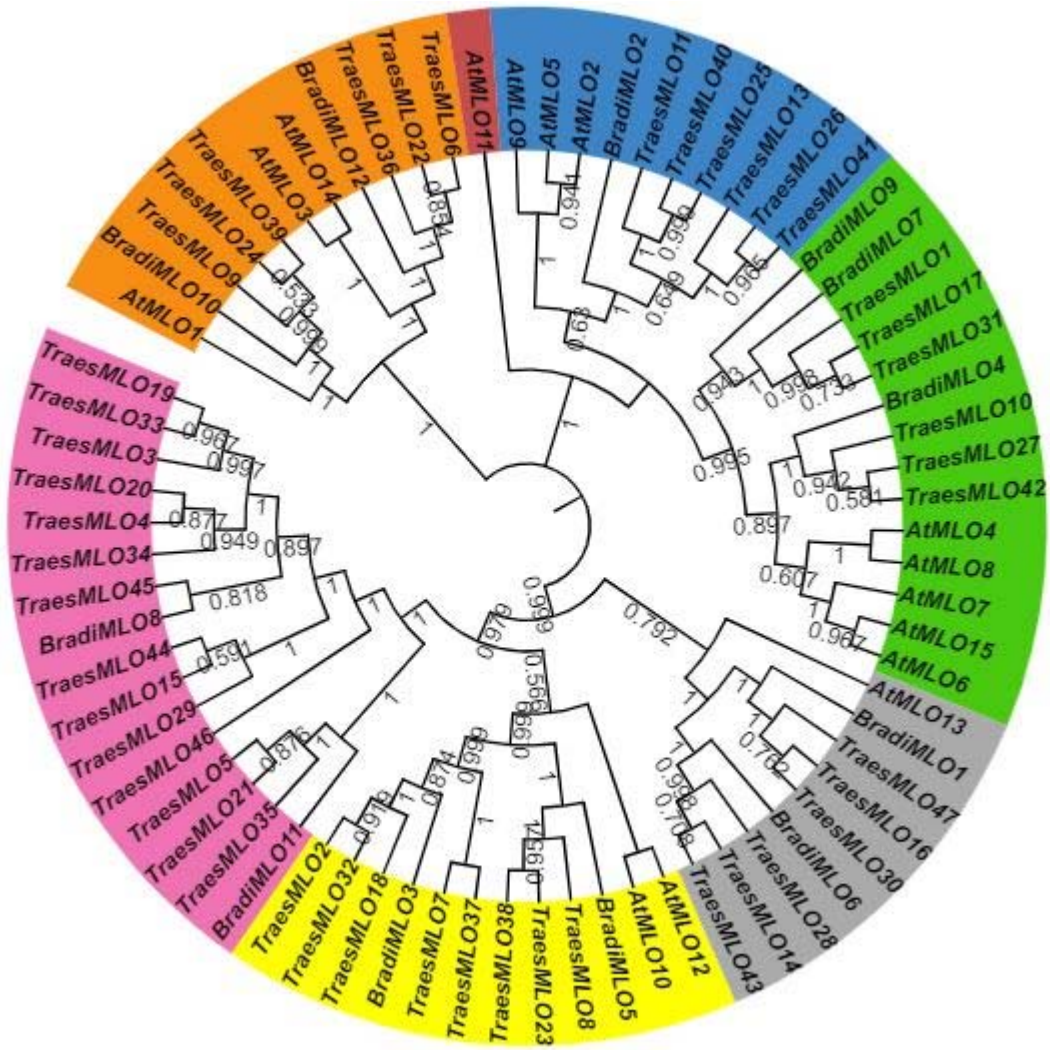

Figure 4: Comparative phylogenetic tree of 47 *TraesMLO* genes with *Arabidopsis thaliana* and *brachypodium distachyon* proteins

3.6. Gene Duplication and Evolution Analysis of wheat and *Arabidopsis thaliana* MLOs

The coding sequences of all TraesMLO and AtMLO genes were used for gene duplication analysis. The gene pairs/alignments with ≥90% sequence homology were considered to duplicated during process of evolution. Total 50 duplicated gene pairs in bread wheat passed this criterion and corresponded to 42 TraesMLO genes. Among these 50 duplications, 47 were segmental i.e., duplicated genes occurred on different chromosomes and remaining three were tandem repeats i.e., duplicated pairs occurred in close vicinity on the same chromosome. These three tandem duplications were found on Chr 2A, 2B, and 2D (**Table 5**). Thus segmental duplications were the major route of evolution for wheat MLOs that has also been reported for MADS-Box gene family in bread wheat (Raza et al., 2022), and APETALA-2 gene family in rice (Ahmed et al., 2021).

Further evolution analysis was performed by calculating the synonymous substitution rate (Ks), and non-synonymous substitution rate (Ka) per site per year of all duplicated MLO gene pairs. These were used to calculate the Ka/Ks ratios to measure the selection pressure on amino-acid substitutions during the evolution. The values of Ka/Ks ranged between 0.052 and 0.373, as Ka/Ks < 1, it implied that synonymous substitutions were more prevalent as compared to non-synonymous changes (**Table 5**). Therefore, all duplicated genes had negative or purifying selection type that aims to remove deleterious alleles during the evolution process. Such purifying changes are quite common in wheat (Raza et al., 2022) and rice as well (Ahmed et al., 2021). We used the values of Ks to calculate the time of divergence/duplication and it was found to be between 3.60 and 19.83 million years ago (MYA). However, we did not find any duplication event in AtMLO genes at 90% identity criteria. However, five duplications were found when identity score was reduced to 70%. Among them 80% were segmental (4 out of 5) and had purifying selection effects just like wheat (**Table 6**). In this way, evolution analysis of AtMLOs supported the wheat. However, their divergence times were quite high as compared to bread wheat.

Table 5: Gene duplication and evolution analysis of MLO genes and their estimated divergence time (Million Years Ago, MYO) in bread wheat

S#	Gene 1	Gene 2	Chr.*(Gene 1; Gene 2)	% Identity	Duplication Type	Ks	Kn	Kn/Ks	Selection Type	Estimated Divergence Time (MYA)
1	*TraesMLO40*	*TraesMLO25*	4D; 4B	96.30	Segmental	0.080	0.026	0.323	Purifying	6.14
2	*TraesMLO13*	*TraesMLO26*	5A; 4B	95.00	Segmental	0.173	0.020	0.116	Purifying	13.28
3	*TraesMLO13*	*TraesMLO41*	5A; 4D	95.64	Segmental	0.136	0.020	0.148	Purifying	10.42
4	*TraesMLO26*	*TraesMLO41*	4B; 4D	97.13	Segmental	0.102	0.009	0.087	Purifying	7.87
5	*TraesMLO1*	*TraesMLO31*	1A; 1B	97.56	Segmental	0.088	0.007	0.079	Purifying	6.78
6	*TraesMLO1*	*TraesMLO17*	1A; 1D	97.27	Segmental	0.106	0.006	0.058	Purifying	8.17

7	*TraesMLO31*	*TraesMLO17*	1B; 1D	98.40	Segmental	0.051	0.006	0.121	Purifying	3.95
8	*TraesMLO27*	*TraesMLO42*	5B; 5D	98.00	Segmental	0.058	0.011	0.286	Purifying	3.60
9	*TraesMLO27*	*TraesMLO10*	5B; 5A	97.83	Segmental	0.047	0.013	0.170	Purifying	5.61
10	*TraesMLO42*	*TraesMLO10*	5D; 5A	96.78	Segmental	0.073	0.012	0.170	Purifying	5.61
11	*TraesMLO43*	*TraesMLO14*	6D; 6A	96.99	Segmental	0.097	0.012	0.125	Purifying	7.48
12	*TraesMLO43*	*TraesMLO28*	6D; 6B	97.31	Segmental	0.084	0.011	0.135	Purifying	6.48
13	*TraesMLO14*	*TraesMLO28*	6A; 6B	97.79	Segmental	0.068	0.010	0.141	Purifying	5.25
14	*TraesMLO30*	*TraesMLO16*	7B; 7A	96.36	Segmental	0.091	0.024	0.261	Purifying	6.99
15	*TraesMLO30*	*TraesMLO47*	7B; 7D	95.18	Segmental	0.113	0.033	0.288	Purifying	8.67
16	*TraesMLO16*	*TraesMLO47*	7A; 7D	95.82	Segmental	0.081	0.030	0.373	Purifying	6.25
17	*TraesMLO8*	*TraesMLO23*	3A; 3B	95.00	Segmental	0.149	0.025	0.170	Purifying	11.43
18	*TraesMLO8*	*TraesMLO38*	3A; 3D	94.48	Segmental	0.171	0.026	0.152	Purifying	13.17
19	*TraesMLO23*	*TraesMLO38*	3B; 3D	97.73	Segmental	0.072	0.009	0.124	Purifying	5.56
20	*TraesMLO7*	*TraesMLO37*	3B; 3D	97.45	Segmental	0.061	0.017	0.269	Purifying	4.72
21	*TraesMLO18*	*TraesMLO2*	1B; 1A	91.87	Segmental	0.193	0.057	0.297	Purifying	14.82
22	*TraesMLO18*	*TraesMLO32*	1B; 1D	92.00	Segmental	0.191	0.056	0.292	Purifying	14.72
23	*TraesMLO2*	*TraesMLO32*	1A; 1D	96.93	Segmental	0.125	0.007	0.052	Purifying	9.63
24	*TraesMLO5*	*TraesMLO35*	2A; 2D	96.07	Segmental	0.112	0.019	0.172	Purifying	8.60
25	*TraesMLO5*	*TraesMLO21*	2A; 2B	95.64	Segmental	0.144	0.018	0.123	Purifying	11.07
26	*TraesMLO35*	*TraesMLO21*	2D; 2B	96.81	Segmental	0.092	0.016	0.172	Purifying	7.05
27	*TraesMLO44*	*TraesMLO15*	6D; 6A	98.38	Segmental	0.058	0.005	0.089	Purifying	4.48
28	*TraesMLO44*	*TraesMLO29*	6D; 6B	97.72	Segmental	0.070	0.011	0.149	Purifying	5.41
29	*TraesMLO15*	*TraesMLO29*	6A; 6B	97.02	Segmental	0.081	0.017	0.207	Purifying	6.22
30	*TraesMLO3*	*TraesMLO19*	2A; 2B	95.20	Segmental	0.157	0.021	0.135	Purifying	12.04
31	*TraesMLO3*	*TraesMLO33*	2A; 2D	96.35	Segmental	0.127	0.013	0.104	Purifying	9.76
32	*TraesMLO19*	*TraesMLO33*	2B; 2D	97.77	Segmental	0.076	0.008	0.104	Purifying	5.84
33	*TraesMLO3*	*TraesMLO34*	2A; 2D	92.82	Segmental	0.254	0.031	0.120	Purifying	19.53
34	*TraesMLO3*	*TraesMLO4*	2A; 2A	93.24	Tandem	0.206	0.035	0.169	Purifying	15.85
35	*TraesMLO3*	*TraesMLO20*	2A; 2B	92.90	Segmental	0.236	0.034	0.142	Purifying	18.18
36	*TraesMLO19*	*TraesMLO34*	2B; 2D	92.61	Segmental	0.258	0.033	0.127	Purifying	19.83
37	*TraesMLO19*	*TraesMLO4*	2B; 2A	93.17	Segmental	0.212	0.035	0.165	Purifying	16.27
38	*TraesMLO19*	*TraesMLO20*	2B; 2B	93.31	Tandem	0.213	0.034	0.157	Purifying	16.40
39	*TraesMLO33*	*TraesMLO34*	2D; 2D	93.89	Tandem	0.206	0.028	0.137	Purifying	15.87
40	*TraesMLO33*	*TraesMLO4*	2D; 2A	94.33	Segmental	0.164	0.030	0.185	Purifying	12.61
41	*TraesMLO33*	*TraesMLO20*	2D; 2B	94.06	Segmental	0.181	0.031	0.170	Purifying	13.91
42	*TraesMLO34*	*TraesMLO4*	2D; 2A	96.03	Segmental	0.146	0.012	0.082	Purifying	11.19
43	*TraesMLO34*	*TraesMLO20*	2D; 2B	96.16	Segmental	0.144	0.011	0.079	Purifying	11.04
44	*TraesMLO4*	*TraesMLO20*	2A; 2B	97.34	Segmental	0.086	0.010	0.113	Purifying	6.65
45	*TraesMLO9*	*TraesMLO24*	4A; 4B	97.83	Segmental	0.071	0.009	0.122	Purifying	5.43
46	*TraesMLO9*	*TraesMLO39*	4A; 4D	97.57	Segmental	0.078	0.009	0.119	Purifying	5.97
47	*TraesMLO24*	*TraesMLO39*	4B; 4D	98.27	Segmental	0.059	0.006	0.104	Purifying	4.51
48	*TraesMLO6*	*TraesMLO36*	2A; 2D	95.79	Segmental	0.122	0.016	0.134	Purifying	9.40
49	*TraesMLO6*	*TraesMLO22*	2A; 2B	94.83	Segmental	0.181	0.019	0.105	Purifying	13.93

50	*TraesMLO36*	*TraesMLO22*	2D; 2B	96.95	Segmental	0.106	0.010	0.094	Purifying	8.16

Table 6: Gene duplication and evolution analysis of MLO genes and their estimated divergence time (Million Years Ago, MYO) in *Arabidopsis thaliana*

S#	Gene 1	Gene 2	Chr.*(Gene 1; Gene 2)	% Identity	Duplication Type	Ks	Ka	Ka/Ks	Selection Type	Estimated Divergence Time (MYA)
1	*AtMLO9*	*AtMLO2*	Chr2; Chr1	70.94	Segmental	0.481	0.325	0.675	Purifying	37.02
2	*AtMLO9*	*AtMLO5*	Chr2; Chr1	70.22	Segmental	0.519	0.361	0.695	Purifying	39.91
3	*AtMLO2*	*AtMLO5*	Chr1; Chr1	78.56	Tandem	0.344	0.231	0.670	Purifying	26.48
4	*AtMLO6*	*AtMLO15*	Chr2; Chr5	75.78	Segmental	0.420	0.274	0.653	Purifying	32.29
5	*AtMLO14*	*AtMLO3*	Chr5; Chr1	80.49	Segmental	0.910	0.104	0.114	Purifying	69.96

3.7. Expression Analysis of MLO genes

3.7.1. TaMLOs expression at different stages of developments

MLO genes expression patterns have been extensively investigated in crop plants during growth and development, as well as under biotic and abiotic stresses. In this study, we analyzed *TaMLO*s expression perturbations in several tissues at different time points to get insights into functional roles under prevalent conditions. Out of total 47 genes, 28 (~60%) were expressed in at least two different tissues during whole growth and development period (**Figure 5A**). Among these expressed genes, at least 9 (*TaMLO2, TaMLO3, TaMLO13, TaMLO14, TaMLO26, TaMLO28, TaMLO41, TaMLO43* and *TaMLO45*) were ubiquitously expressed in nearly all investigated tissues. Interestingly, all these genes exhibited highly significant expression patterns in flag leaf sheath and blade during reproductive phase (30% spike to grain ripening stages) (**Figure 5A**). However, several other genes exhibited tissue specific and growth stage specific expression perturbations. Collectively, these expression patterns indicate diverse functions of MLO genes in wheat growth and development.

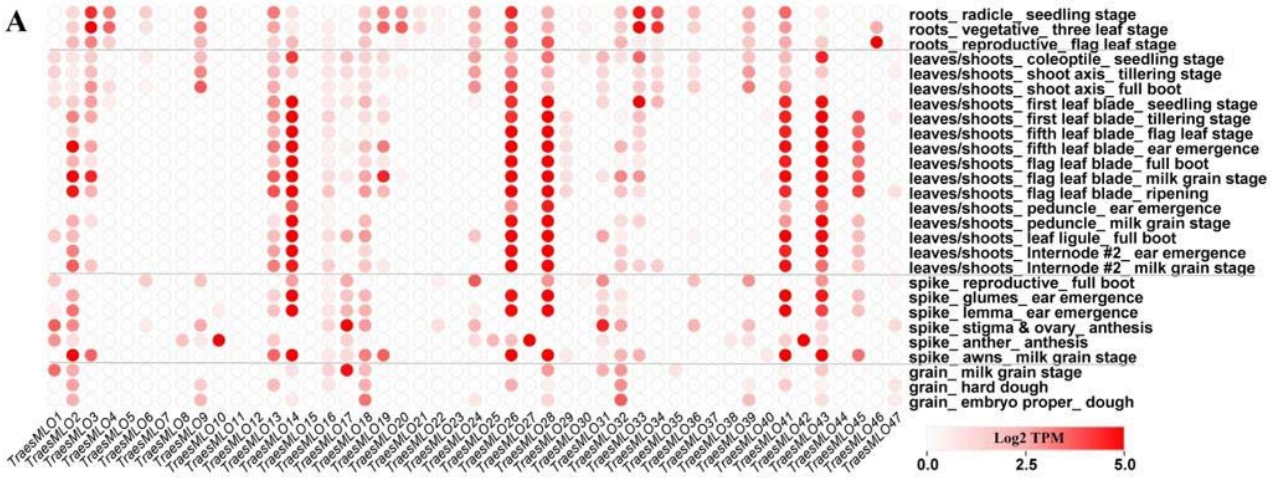

Figure 4A： Expression patterns of TaMLOs during different stages of growth and development.

3.7.2. TaMLOs expression at different biotic and abiotic stresses

We also studied the transcriptional changes in *TaMLO*s were under four different abiotic stresses including cold, drought, heat and phosphate starvation (**Figure 5B**), as well as under three different biotic stresses including fusarium head blight (FHB), powdery mildew (PM) and stripe rust (SR) inoculations (**Figure 5C**). Likewise, expression patterns during growth and development, almost ≥60% genes were expressed after at least one abiotic and/or biotic stress application (**Figure 5B** & **5C**).

Majority of these genes expressed under phosphate starvation and FHB infections. Interestingly, all those genes which were highly ubiquitously expressed during different growth and development stages were also highly expressed under abiotic and biotic stresses, except for *TaMLO45*. Under abiotic stresses, seven genes after two weeks of cold stress (*TaMLO1*, *TaMLO13*, *TaMLO18*, *TaMLO31*, *TaMLO32*, *TaMLO33* and *TaMLO45*), six genes after 1–12 hours of drought stress (*TaMLO4*, *TaMLO19*, *TaMLO20*, *TaMLO32*, *TaMLO33* and *TaMLO34*), three genes after 1–6 hours of heat stress (*TaMLO29*, *TaMLO33* and *TaMLO44*) and three genes under phosphate starvation at vegetative stage (*TaMLO16*, *TaMLO33* and *TaMLO39*) exhibited substantial transcriptional perturbations compared with their respective control (CK) treatments (**Figure 5B**).

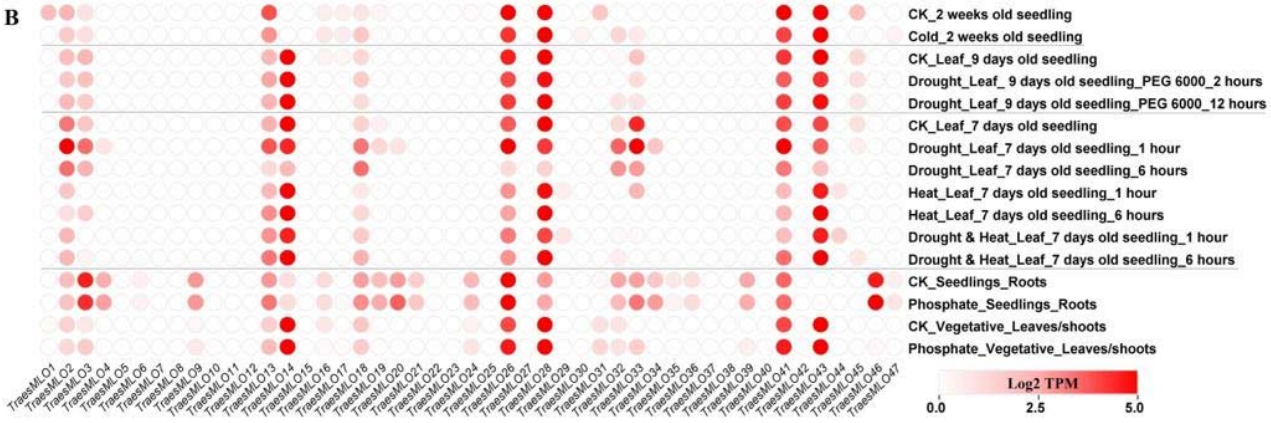

Figure 5B: Expression patterns of TaMLOs under different abiotic stresses.

Under biotic stresses, at least four genes (*TaMLO4, TaMLO13, TaMLO16 and TaMLO46*) exhibited powdery mildew specific transcriptional perturbation when compared with their respective control treatment. Whereas, only *TaMLO32* showed stripe rust specific expression changes. Likewise, at least five genes (*TaMLO4, TaMLO6, TaMLO19, TaMLO20, TaMLO33* and *TaMLO34*) exhibited FHB specific expression patterns when compared with respective control treatments (**Figure 5C; Figure 5S)**. Overall, these transcriptional perturbations during growth and development and under different abiotic and biotic stresses indicate that *TaMLO*s regulate multifarious processes to successfully complete bread wheat life cycle under continuously changing environmental conditions.

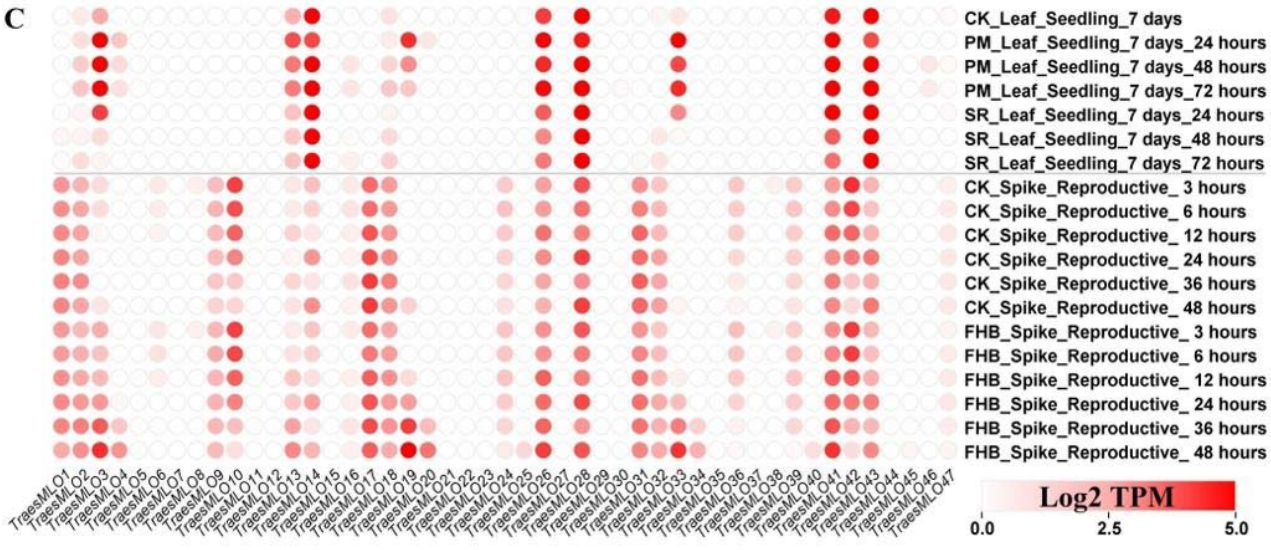

Figure 5C: Expression patterns of TaMLOs under different biotic stresses.

MLO family genes are unique to plants, governing resistance against powdery mildew infections through a negative regulatory pathway However, apart from conferring plant immunity, previous studies have also reported their diverse roles in growth and development and stress responses (Bindzinski et al., 2014; Tian et al., 2022), just like multifarious roles of MADS-box family genes in bread wheat (Raza et al., 2022). For example, a systematic study of rice MLO genes indicated their roles in morphological development, photoperiod sensitivity, abiotic stress tolerance and resistance against different biotic stresses (Nguyen et al., 2016). In this study, a substantial number of MLO genes also showed ubiquitous expression patterns during growth and development, abiotic and biotic stresses (**Figure 5A**, **5B** & **5C**). These results support previous findings about functional roles of MLO genes and reveal that these plant specific genes contribute paramount roles during successful completion of wheat plant life cycle and survival under unfavorable conditions. Detailed functional characterization of significantly perturbed MLO genes using modern molecular and biotechnological techniques such as CRISPR-Cas could yield strong candidates for development of climate-resilient and super-immune wheat cultivars.

References

Ablazov, A., and Tombuloglu, H. (2016). Genome-wide identification of the mildew resistance locus O (MLO) gene family in novel cereal model species Brachypodium distachyon. *Eur. J. Plant Pathol.* 145, 239–253. doi:10.1007/s10658-015-0833-2.

Ahmed, S., Rashid, M. A. R., Zafar, S. A., Azhar, M. T., Waqas, M., Uzair, M., et al. (2021). Genome-wide investigation and expression analysis of APETALA-2 transcription factor subfamily reveals its evolution, expansion and regulatory role in abiotic stress responses in Indica Rice (Oryza sativa L. ssp. indica). *Genomics* 113, 1029–1043. doi:10.1016/j.ygeno.2020.10.037.

Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science (80-.).* 361. doi:10.1126/science.aar7191.

Bindzinski, P., NOIR, S., SHAHI, S., REINSTÄDLER, A., GRATKOWSKA, D. M., and PANSTRUGA, R. (2014). Physiological characterization and genetic modifiers of aberrant root thigmomorphogenesis in mutants of <scp> *A* </scp> *rabidopsis thaliana* <scp> *MILDEW LOCUS O* </scp> genes. *Plant. Cell Environ.* 37, 2738–2753. doi:10.1111/pce.12353.

Borrill, P., Ramirez-Gonzalez, R., and Uauy, C. (2016). expVIP: a Customizable RNA-seq Data Analysis and Visualization Platform. *Plant Physiol.* 170, 2172–2186. doi:10.1104/pp.15.01667.

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* 13, 1194–1202. doi:10.1016/j.molp.2020.06.009.

Chen, L., Liu, J., Liu, Z., Xia, P., Yuan, X., and Ning, Y. (2021). Genome-wide identification and expression analysis of the MLO gene family reveal a candidate gene associated with powdery mildew susceptibility in bitter gourd (Momordica charantia). *Eur. J. Plant Pathol.* 159, 163–178. doi:10.1007/s10658-020-02152-0.

Deshmukh, R., Singh, V. K., and Singh, B. D. (2014). Comparative phylogenetic analysis of genome-wide Mlo gene family members from Glycine max and Arabidopsis thaliana. *Mol. Genet. Genomics* 289, 345–359. doi:10.1007/s00438-014-0811-y.

Devoto, A., Hartmann, H. A., Piffanelli, P., Elliott, C., Simmons, C., Taramino, G., et al. (2003). Molecular phylogeny and evolution of the plant-specific seven-transmembrane MLO family. *J. Mol. Evol.* 56, 77–88. doi:10.1007/s00239-002-2382-5.

El Baidouri, M., Murat, F., Veyssiere, M., Molinier, M., Flores, R., Burlot, L., et al. (2017). Reconciling the evolutionary origin of bread wheat (Triticum aestivum). *New Phytol.* 213, 1477–1486. doi:10.1111/nph.14113.

Gao, L., McCarthy, E. M., Ganko, E. W., and McDonald, J. F. (2004). Evolutionary history of Oryza sativa LTR retrotransposons: A preliminary survey of the rice genome sequences. *BMC Genomics* 5, 1–18. doi:10.1186/1471-2164-5-18.

Hurni, S., Brunner, S., Buchmann, G., Herren, G., Jordan, T., Krukowski, P., et al. (2013). Rye Pm8 and wheat Pm3 are orthologous genes and show evolutionary conservation of resistance function against powdery mildew. *Plant J.* 76, 957–969. doi:10.1111/tpj.12345.

Hussain, B. (2015). Modernization in plant breeding approaches for improvingbiotic stress resistance in crop plants. *TURKISH J. Agric. For.* 39, 515–530. doi:10.3906/tar-1406-176.

Hussain, B., and Ahmad, S. (2022). “CRISPR/Cas9 for Rice Crop Improvement: Recent Progress, Limitations, and Prospects,” in *Modern Techniques of Rice Crop Production* (Singapore: Springer Singapore), 701–717. doi:10.1007/978-981-16-4955-4_34.

Hussain, B., Akpınar, B. A., Alaux, M., Algharib, A. M., Sehgal, D., Ali, Z., et al. (2022).

Capturing Wheat Phenotypes at the Genome Level. *Front. Plant Sci.* 13. doi:10.3389/fpls.2022.851079.

Hussain, B., Lucas, S. J., and Budak, H. (2018). CRISPR/Cas9 in plants: at play in the genome and at work for crop improvement. *Brief. Funct. Genomics*, 1–10. doi:10.1093/bfgp/ely016.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-.).* 337, 816–821. doi:10.1126/science.1225829.

Konishi, S., Sasakuma, T., and Sasanuma, T. (2010). Identification of novel Mlo family members in wheatand their genetic characterization. *Genes Genet. Syst.* 85, 167–175. doi:10.1266/ggs.85.167.

Krattinger, S. G., Lagudah, E. S., Spielmeyer, W., Singh, R. P., Huerta-Espino, J., McFadden, H., et al. (2009). A Putative ABC Transporter Confers Durable Resistance to Multiple Fungal Pathogens in Wheat. *Science (80-.).* 323, 1360–1363. doi:10.1126/science.1166453.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi:10.1093/molbev/msy096.

Kusch, S., Pesch, L., and Panstruga, R. (2016). Comprehensive phylogenetic analysis sheds light on the diversity and origin of the MLO family of integral membrane proteins. *Genome Biol. Evol.* 8, 878–895. doi:10.1093/gbe/evw036.

Li, M., Dong, L., Li, B., Wang, Z., Xie, J., Qiu, D., et al. (2020). A CNL protein in wild emmer wheat confers powdery mildew resistance. *New Phytol.* 228, 1027–1037. doi:10.1111/nph.16761.

Li, S., Lin, D., Zhang, Y., Deng, M., Chen, Y., Lv, B., et al. (2022). Genome-edited powdery mildew resistance in wheat without growth penalties. *Nature* 602, 455–460. doi:10.1038/s41586-022-04395-9.

Lu, P., Guo, L., Wang, Z., Li, B., Li, J., Li, Y., et al. (2020). A rare gain of function mutation in a wheat tandem kinase confers resistance to powdery mildew. *Nat. Commun.* 11, 680. doi:10.1038/s41467-020-14294-0.

McIntosh, R. A., Dubcovsky, J., Rogers, W. J., Xia, X. C., and Raupp, W. J. CATALOGUE OF GENE SYMBOLS FOR WHEAT: 2019 SUPPLEMENT. *A n n u a l W h e a t N e ws Let t e r*. Available at: https://wheat.pw.usda.gov/GG3/sites/default/files/Catalogue of Gene Symbols

for Wheat - supplement2019.pdf.

Moore, J. W., Herrera-Foessel, S., Lan, C., Schnippenkoetter, W., Ayliffe, M., Huerta-Espino, J., et al. (2015). A recently evolved hexose transporter variant confers resistance to multiple pathogens in wheat. *Nat. Genet.* 47, 1494–1498. doi:10.1038/ng.3439.

Muhire, B. M., Varsani, A., and Martin, D. P. (2014). SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. *PLoS One* 9, e108277. doi:10.1371/journal.pone.0108277.

Nguyen, V. N. T., Vo, K. T. X., Park, H., Jeon, J.-S., and Jung, K.-H. (2016). A Systematic View of the MLO Family in Rice Suggests Their Novel Roles in Morphological Development, Diurnal Responses, the Light-Signaling Pathway, and Various Stress Responses. *Front. Plant Sci.* 7. doi:10.3389/fpls.2016.01413.

Panchy, N., Lehti-Shiu, M., and Shiu, S. H. (2016). Evolution of gene duplication in plants. *Plant Physiol.* 171, 2294–2316. doi:10.1104/pp.16.00523.

Polanco, C., Sáenz De Miera, L. E., Bett, K., and De La Vega, M. P. (2018). A genome-wide identification and comparative analysis of the lentil MLO genes. *PLoS One* 13, e0194945. doi:10.1371/journal.pone.0194945.

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., and Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Res.* 46, W200–W204. doi:10.1093/nar/gky448.

Ramírez-González, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., et al. (2018). The transcriptional landscape of polyploid wheat. *Science (80-.).* 361. doi:10.1126/science.aar6089.

Raza, Q., Riaz, A., Atif, R. M., Hussain, B., Rana, I. A., Ali, Z., et al. (2022). Genome-Wide Diversity of MADS-Box Genes in Bread Wheat is Associated with its Rapid Global Adaptability. *Front. Genet.* 12, 2802. doi:10.3389/fgene.2021.818880.

Sánchez-Martín, J., Steuernagel, B., Ghosh, S., Herren, G., Hurni, S., Adamski, N., et al. (2016). Rapid gene isolation in barley and wheat by mutant chromosome sequencing. *Genome Biol.* 17, 221. doi:10.1186/s13059-016-1082-1.

Scholthof, K. B. G. (2007). The disease triangle: Pathogens, the environment and society. *Nat. Rev. Microbiol.* 5, 152–156. doi:10.1038/nrmicro1596.

Shan, Q., Wang, Y., Li, J., and Gao, C. (2014). Genome editing in rice and wheat using the CRISPR/Cas system. *Nat. Protoc.* 9, 2395–2410. doi:10.1038/nprot.2014.157.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539. doi:10.1038/msb.2011.75.

Singh, R. P., Singh, P. K., Rutkoski, J., Hodson, D. P., He, X., Jørgensen, L. N., et al. (2016). Disease Impact on Wheat Yield Potential and Prospects of Genetic Control. *Annu. Rev. Phytopathol.* 54, 303–322. doi:10.1146/annurev-phyto-080615-095835.

Singh, V. K., Singh A K, Chand R, and Singh B D (2012). GENOME WIDE ANALYSIS OF DISEASE RESISTANCE MLO GENE FAMILY IN SORGHUM [Sorghum bicolor (L.) MOENCH] Journal of Plant Genomics. *J. Plant Genomics* 2, 18–27.

Tapia, R. R., Barbey, C. R., Chandra, S., Folta, K. M., Whitaker, V. M., and Lee, S. (2021). Evolution of the MLO gene families in octoploid strawberry (Fragaria × ananassa) and progenitor diploid species identified potential genes for strawberry powdery mildew resistance. *Hortic. Res.* 8. doi:10.1038/s41438-021-00587-y.

Tian, Y., Wang, Q., Zhang, H., Zhou, N., Yan, H., Jian, H., et al. (2022). Genome-wide identification and evolutionary analysis of MLO gene family in Rosaceae plants. *Hortic. Plant J.* 8, 110–122. doi:10.1016/j.hpj.2020.07.007.

Wang, J., Wu, X., Wang, Y., Wu, X., Wang, B., Lu, Z., et al. (2023). Genome-wide characterization and expression analysis of the MLO gene family sheds light on powdery mildew resistance in Lagenaria siceraria. *Heliyon* 9, e14624. doi:10.1016/j.heliyon.2023.e14624.

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., et al. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* 32, 947–951. doi:10.1038/nbt.2969.

Win, K. T., Zhang, C., and Lee, S. (2018). Genome-wide identification and description of MLO family genes in pumpkin (Cucurbita maxima Duch.). *Hortic. Environ. Biotechnol.* 59, 397–410. doi:10.1007/s13580-018-0036-9.

Xie, J., Guo, G., Wang, Y., Hu, T., Wang, L., Li, J., et al. (2020). A rare single nucleotide variant in Pm5e confers powdery mildew resistance in common wheat. *New Phytol.* 228, 1011–1026. doi:10.1111/nph.16762.

Xing, L., Hu, P., Liu, J., Witek, K., Zhou, S., Xu, J., et al. (2018). Pm21 from Haynaldia villosa Encodes a CC-NBS-LRR Protein Conferring Powdery Mildew Resistance in Wheat. *Mol.*

Plant 11, 874–878. doi:10.1016/j.molp.2018.02.013.

Xu, G., Guo, C., Shan, H., and Kong, H. (2012). Divergence of duplicate genes in exon–intron structure. *Proc. Natl. Acad. Sci.* 109, 1187–1192.

Yahiaoui, N., Srichumpa, P., Dudler, R., and Keller, B. (2004). Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene Pm3b from hexaploid wheat. *Plant J.* 37, 528–538. doi:10.1046/j.1365-313X.2003.01977.x.

Zhang, T., Xu, N., Amanullah, S., and Gao, P. (2023). Genome-wide identification, evolution, and expression analysis of MLO gene family in melon (Cucumis melo L.). *Front. Plant Sci.* 14, 1–15. doi:10.3389/fpls.2023.1144317.

Zhu, T., Wang, L., Rimbert, H., Rodriguez, J. C., Deal, K. R., De Oliveira, R., et al. (2021). Optical maps refine the bread wheat Triticum aestivum cv. Chinese Spring genome assembly. *Plant J.* 107, 303–314. doi:10.1111/tpj.15289.

Zou, S., Wang, H., Li, Y., Kong, Z., and Tang, D. (2018). The NB-LRR gene Pm60 confers powdery mildew resistance in wheat. *New Phytol.* 218, 298–309. doi:10.1111/nph.14964.