

Comparative phylogenetic analysis of genome-wide *Mlo* gene family members from *Glycine max* and *Arabidopsis thaliana*

Reena Deshmukh · V. K. Singh · B. D. Singh

Received: 23 May 2013 / Accepted: 3 January 2014 / Published online: 28 January 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract Powdery mildew locus *O* (*Mlo*) gene family is one of the largest seven transmembrane protein-encoding gene families. The *Mlo* proteins act as negative regulators of powdery mildew resistance and a loss-of-function mutation in *Mlo* is known to confer broad-spectrum resistance to powdery mildew. In addition, the *Mlo* gene family members are known to participate in various developmental and biotic and abiotic stress response-related pathways. Therefore, a genome-wide similarity search using the characterized *Mlo* protein sequences of *Arabidopsis thaliana* was carried out to identify putative *Mlo* genes in soybean (*Glycine max*) genome. This search identified 39 *Mlo* domain containing protein-encoding genes that were distributed on 15 of the 20 *G. max* chromosomes. The putative promoter regions of these *Mlo* genes contained response elements for different external stimuli, including different hormones and abiotic stresses. Of the 39 Gm*Mlo* proteins, 35 were rich (8.7–13.1 %) in leucine, while five were serine-rich (9.2–11.9 %). Furthermore, all the Gm*Mlo* members were localized in the plasma membrane. Phylogenetic analysis of the Gm*Mlo* and the At*Mlo* proteins classified them into three

main clusters, and the cluster I comprised two sub-clusters. Multiple sequence alignment visualized the location of seven transmembrane domains, and a conserved CaM-binding domain. Some of the Gm*Mlo* proteins (Gm*Mlo*10, 20, 22, 23, 32, 36, 37) contained less than seven transmembrane domains. The motif analysis yielded 27 motifs; out of these, motif 2, the only motif present in all the Gm*Mlos*, was highly conserved and three amino acid residues were essentially invariant. Five of the Gm*Mlo* members were much smaller in size; presumably they originated through deletion following a gene duplication event. The presence of a large number of Gm*Mlo* members in the *G. max* genome may be due to its paleopolyploid nature and the large genome size as compared to that of *Arabidopsis*. The findings of this study may further help in characterization and isolation of individual Gm*Mlo* members.

Keywords Comparative phylogeny · *Glycine max* · *Mlo* gene family · Powdery mildew · *Arabidopsis thaliana*

Introduction

Powdery mildew is one of the most ubiquitous plant diseases and it infects a wide variety of plant species, including cereals, millets, pulses, horticultural crops and forest trees. The powdery mildew fungus, *Microsphaera diffusa* Cke. & Pk, is reported to cause yield losses of 0–26 % (Dunleavy 1980) in soybean. The powdery mildew locus *O* (*Mlo*) gene was first identified in barley as being involved in modulation of the defense response against the powdery mildew pathogen, *Blumeria graminis* f.sp. *hordei* (Jørgensen 1992; Büschges et al. 1997). It was later observed that complete resistance to this fungal pathogen was conferred by the mutant allele, *mlo*, of this gene when it was

Communicated by Y. Van de Peer.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-014-0811-y) contains supplementary material, which is available to authorized users.

R. Deshmukh · B. D. Singh (✉)
Faculty of Science, School of Biotechnology,
Banaras Hindu University, Varanasi 221005, India
e-mail: brahmadsingh@gmail.com

V. K. Singh
Faculty of Science, School of Biotechnology, Centre
of Bioinformatics, Banaras Hindu University,
Varanasi 221005, India

present in homozygous state. The resistance conferred by *mlo* allele is broad-spectrum, covering all the strains of the pathogen, and is durable in nature (Jørgensen 1992; Piffanelli et al. 2002). A similar broad-spectrum resistance to powdery mildew due to the loss of *Mlo* function has been reported in tomato (Bai et al. 2008). Thus, *Mlo* function appears to be necessary for successful infection by the powdery mildew pathogen; in fact, the *Mlo* proteins are absolute necessity for this pathogen to be able to successfully penetrate the host cell wall (Piffanelli et al. 2002). This conclusion is further supported by the observation that an over-expression of the wild-type allele, *Mlo*, of this gene leads to super-susceptibility to the powdery mildew pathogen (Wolter et al. 1993; Kim et al. 2002b). There is some evidence that the powdery mildew resistance in common bean (*Phaseolus vulgaris*) is a quantitatively governed character (Trabanco et al. 2012), indicating that some other gene products are also involved in modulation of host response to the powdery mildew pathogen.

Although the *Mlo* gene was originally described for its role in defense response, there were indications that it might also be involved in other biological processes. The homozygous *mlo* plants were reported to show spontaneous death of mesophyll cells associated with increased leaf senescence (Wolter et al. 1993; Kumar et al. 2001; Piffanelli et al. 2002). In addition, *Mlo* gene family seems to be involved in modulation of response to other biotic and abiotic stresses, including wounding, the carbohydrate elicitor produced by wheat powdery mildew, salt stress, mannitol treatment (Piffanelli et al. 2002; Konishi et al. 2010), etc. In relation to the defense response, *Mlo* proteins are commonly expressed during invasion by fungal haustoria/appressoria (Kumar et al. 2001). However, these proteins are constitutively expressed in developmental functions and are also involved in response mechanisms regulated by biotic and abiotic stimuli (Chen et al. 2006). The *Mlo* proteins have a characteristic feature of seven transmembrane (TM) domains (Devoto et al. 1999), and are located in the plasma membrane. These proteins also have a common feature of 20 amino acid long CaM-binding domain (CaMBD; Kim et al. 2002a, b), which mediates the defense signaling mechanism (Reddy et al. 2003). The *Mlo* proteins invariably contain cysteine and proline residues either in their extracellular loops or in their TM domains; these amino acid residues have remained unchanged for over 400 million years and they are essential for *Mlo* protein function and/or stability (Elliott et al. 2005). The second and third cytoplasmic loops of *Mlo* proteins play a critical role in powdery mildew susceptibility (Reinstädler et al. 2010).

The *Mlo* genes are postulated to trace back to the early stages of evolution of land plants, and they are reported to be restricted to higher plants and certain mosses (Devoto et al. 1999, 2003). A large family of RAC/ROP GTPases acts against pathogens in defense responses, and in cytosolic

signal transduction involved in several developmental and stress response phenomena (Chen et al. 2010). The *Mlo* proteins are believed to participate in signal transduction by altering cell polarity in conjunction with RAC G-proteins leading to resistance to powdery mildew fungus in barley (Schultheiss et al. 2002; Opalski et al. 2005). Nibau et al. (2006) have reviewed the role of RAC G-proteins in several other cellular metabolic pathways and downstream signaling. The RAC/ROP (rho of plants) proteins act as constitutive signaling molecules regulating fungal invasion and defense response regulated by ROP GTPases (Hoeffle et al. 2011; Huesmann et al. 2012). The *Mlo* proteins are rich in leucine, which comprises 9–13 % of the total amino acid residues (Singh et al. 2012). It has been concluded that proteins rich in leucine residues contribute to defense response mechanisms (Buchanan and Gay 1996; Kędzierski et al. 2004; Torii 2004; Jung et al. 2004). Leucine-rich receptor kinases play a vital role in signal transduction pathways through different protein–protein interactions (Li and Chory 1997; Forsthoeft et al. 2005; Kemmerling et al. 2007), and are involved in downstream defense-regulated pathways and development-related mechanisms (Osakabe et al. 2005).

The availability of complete genome sequence of *Arabidopsis thaliana* facilitated the identification of 15 *A. thaliana Mlo* (*AtMlo*) genes distributed on all the five chromosomes (Devoto et al. 2003). Subsequently, putative *Mlo* genes were described in important cereal crops like rice (11 genes; Liu and Zhu 2008), sorghum (13 genes; Singh et al. 2012) and wheat (7 genes; Konishi et al. 2010) based on bioinformatics analysis of genome sequences. Shen et al. (2012) surveyed the genome sequence database of soybean (<http://www.phytozome.net/soybean.php>, phytozome, Release v4.0; Schmutz et al. 2010) for the *MLO* domain, and identified 20 *Glycine max* L. *Mlo* (*GmMlo*) genes distributed on 13 of the 20 chromosomes. In the Phytozome database, the genome sequence of *G. max* comprises 975 Mb distributed in 20 chromosomes. In comparison, the soybean genome sequence in the NCBI database contains 1,115 Mb organized in 20 chromosomes (<http://www.ncbi.nlm.nih.gov/genome/?term=glycine+max>, 30/10/13). In view of the above, it was suspected that there may be some more members of the *Mlo* gene family in the soybean genome in addition to those described by Shen et al. (2012). Therefore, we searched the NCBI database for *Mlo* gene family members and were able to detect 19 novel *GmMlo* members in addition to the 20 *GmMlo* genes already reported by Shen et al. (2012). These 39 *GmMlo* genes were distributed on 15 of the 20 chromosomes of *G. max*. The *GmMlo* members were analyzed for chromosomal location, gene structure and organization. In addition, a comparative phylogenetic analysis of *AtMlo* and *GmMlo* members was carried out to deduce their orthologous clustering. We also characterized the promoter regions of the 39 *GmMlo* genes

to decipher their putative functions, which may be helpful in designing of experiments for their further characterization, identification and isolation.

Materials and methods

Identification of *Mlo* gene family members in *G. max* genome and their functional characterization

In view of the availability of complete structural and functional characteristics of the 15 *AtMlo* genes, the sequences of proteins encoded by them were used for the identification of *Mlo* gene family members in soybean genome. The amino acid sequences of the *AtMlo* proteins were extracted from TAIR (The *Arabidopsis* Information Resource; <http://www.arabidopsis.org/>) and used as query sequences in the tBLASTn tool (Altschul et al. 1990) for homology search of *GmMlos* from the NCBI soybean genome database (<http://www.ncbi.nlm.nih.gov/sites/entrez>). The homology search was used at 50 % or more query coverage and greater than 35 % identity. The superfluous hits were carefully removed as follows. First, the NCBI accession numbers and the chromosomal locations of the putative *GmMlos* were compared. In case two or more *GmMlos* had the same NCBI accession number and/or the chromosomal location, their amino acid sequences were compared. In case the amino acid sequences also matched, only one of these putative *GmMlos* was retained. Pfam (pfam.sanger.ac.uk/; Punta et al., 2012), the HMMER-based database, and InterProScan (Quevillon et al. 2005), which searches against secondary database, were further used to confirm the presence of signature MLO domain in the putative *GmMlo* sequences.

In order to investigate the probable roles of *GmMlo* members in plant cellular functions, their 1,000-bp upstream sequences representing the putative promoter regions were analyzed for the presence of various *cis*-acting regulatory elements. These upstream regions were analyzed using PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>; Lescot et al. 2002) and PLACE (<http://www.dna.affrc.go.jp/PLACE/>; Higo et al. 1999) tools, which detect and identify the various *cis*-regulatory sequences that act as binding sites for regulatory proteins involved in the regulation of gene expression.

Topological configuration of *GmMlo* proteins and their gene organization

Prediction of TM helices of the putative *GmMlo* proteins was done using HMMTOP server 2.0 (<http://www.enzim.hu/hmmtop/>; Tusnády and Simon 1998, 2001), which provides the localization of proteins as well as the topological arrangement of the TM domains. Sub-cellular localization

of all the *GmMlo* proteins was determined by the CELLO v.2.5 support vector machine software (<http://cello.life.nctu.edu.tw/>; Yu et al. 2006), and their physico-chemical properties were studied using the ProtParam server (<http://web.expasy.org/protparam/>; Gasteiger et al. 2005). The gene structure was studied using the two HMM-based fastest and most accurate gene prediction tools: Fgenesh (<http://linux1.softberry.com/>; Solovyev et al. 2006) and GENSCAN (<http://genes.mit.edu/GENSCAN.html>; Burge and Karlin. 1997). The number of exons was determined from the GNOMON tool as used in the NCBI soybean database. The exon–intron gene organization of the 39 *GmMlo* genes was determined using Gene Structure Display Server (GSDS; <http://gsds.cbi.pku.edu.cn/>; Guo et al. 2007).

Construction of phylogenetic trees and motif study

The multiple sequence alignment of amino acids using ClustalW (Thompson et al. 1994) with default parameters was visualized with CLC Sequence Viewer 6.8.1 software (<http://www.clcbio.com>), which enables multitasking working environment with clear visualization of each amino acid in a different color. The alignment generated through ClustalW was fed to CLC sequence viewer to generate sequence alignment of the 39 *GmMlo* proteins. We used MEGA 5.0 tool to construct a phylogenetic tree based on Maximum likelihood estimation (MLE) and Neighbor-joining (NJ) methods with 1,000 bootstrap replications. The MEME (<http://meme.nbcr.net/>; Bailey and Elkan. 1994, Bailey et al. 2009) package was used to analyze the functional motifs of the *GmMlo* proteins. Correlation of phylogenetic tree with the alignment file was done using Geneious pro 3.5.6 (<http://www.geneious.com/>; Drummond et al. 2008) to draw simultaneous inferences from phylogenetic tree and conserved motifs, and to have a generalized conclusion of motif variation and member diversification. The phylogenetic tree in Geneious pro was constructed using Neighbor-joining method with 1,000 bootstrap replications based on Jukes–Cantor genetic distance model (the only model that can be used in Geneious pro). Synteny between the selected pairs of chromosomal regions containing groups of *GmMlo* genes was analyzed using multiple-genome comparison and alignment tool (MGCAT; <http://algggen.lsi.upc.es/recerca/align/mgcat/>; Treangen and Messegueur. 2006).

Results

Mlo gene identification in *G. max* genome and analysis of their *cis*-acting elements

The amino acid sequences of the 15 *AtMlo* proteins were independently used in BLAST search to trap every

possible homologous *GmMlo* member present in the soybean genome sequence available in the NCBI database. A total of 39 *GmMlo* members were identified by genome-wide scanning using BLAST Entrez query. The nucleotide sequences corresponding to the predicted GmMlo proteins were obtained using tBLASTn. The predicted *GmMlo* genes were numbered 1–20 based on their correspondence with the members of *GmMlo* family reported by Shen et al. (2012); the remaining 19 genes were numbered sequentially from 21 to 39, depending on their chromosomal location (Table 1; Supplementary Table 1). The 39 *GmMlo* genes were present in 15 of the 20 *G. max* chromosomes, the chromosomes devoid of *GmMlo* members were 5, 7, 14, 17 and 18. Chromosome 12 had the maximum number of six *Mlo* genes, followed by chromosome 6 (5 genes); chromosomes 8, 9, 10, 19 and 20 had one gene each, while the remaining eight chromosomes carried between two and four genes (Fig. 1). The two or more *GmMlo* genes located in a single chromosome were generally scattered, but six gene-pairs and two groups of three genes were located rather close to each other and appeared to form clusters. The distances between the presumed transcribed regions of pairs of these genes ranged from 11,725 bp between *GmMlo8* and 27 to 709,271 bp between *GmMlo11* and 29 (Supplementary Table 1).

The analysis of 1,000-bp upstream regions of the *GmMlo* genes indicated the presence of elements involved in hormonal response (ABRE, AuxRR-core, ERE, GARE-motif, P-box, TATC-box, TGA-element, TCA element, GCTCA element and TGACG element), tissue-specific gene expression (shoot-specific: as-2-box; seed-specific: RY-element, GCN4_motif and Skn-1 motif), developmental regulation (circadian, HD-Zip, CCAAT-box and Nodule-site 2 elements), elicitor response (AT-rich region and Box-W1 elements), and defense and abiotic stress responses (TC-rich repeats; Supplementary Tables 2a–2d). In addition, elements involved in wound response (WUN motif), and response to abiotic factors like light (MRE element), anoxia (ARE and GC motif elements), heat shock (HSE element), low temperature (LTR element), and drought (MBS element), were also present. Thus, 32 of the 39 *GmMlo* genes had at least one hormone response element, and two of the genes (*GmMlo15* and *GmMlo24*) had four different hormone response elements each. Similarly, 22 genes had an element for seed-specific expression, and *GmMlo21* had all the three seed-specific elements. Thirteen genes had pathogen elicitor response elements, and elements for developmental roles were present in a similar number of *GmMlos*. Finally, 33 of the *GmMlos* had response element for one or the other abiotic stress; most of these genes had more than one type of response element, and *GmMlo8* and *GmMlo39* had five different response elements each. Among the different motifs, the seed-specific Skn-1 motif was present

in 19 genes; AT-rich repeat motif was found in 17 genes, while HD-Zip, Nodule-site 2, AuxRR-core, P-box and GC motif were detected in only one gene each.

Mlo protein domain characterization and gene organization

The TM helices, their locations in the polypeptide, and the topological arrangement of the GmMlo proteins were studied using HMMTOP server 2.0. The number of TM domains varied from four in *GmMlo37* to up to ten in *GmMlo1*, *GmMlo28*, *GmMlo11* and *GmMlo17*. However, 18 of the 39 GmMlo proteins had seven TM domains, the characteristic number of TM for these proteins (Table 1). All the GmMlo proteins were localized in plasma membrane of the cell (Kim and Hwang 2012) with their N-terminus projecting outward. The ProtParam study of physical and chemical properties of GmMlo proteins indicated that 35 of these proteins were rich in leucine, which comprised from 8.7 % (*GmMlo8*) to 13.1 % (*GmMlo26*) of the total amino acid residues; in 25 of these proteins the leucine content was 10 % or higher. The proteins *GmMlo9*, *GmMlo25*, *GmMlo29*, *GmMlo32* and *GmMlo33*, on the other hand, were serine-rich, and they showed 9.2 % (*GmMlo33*)–11.9 % (*GmMlo9*) serine content. The molecular weights of the GmMlo proteins ranged from 33,325.5 to 78,850.5 Da, their theoretical pI varied from 6.87 to 9.66, aliphatic indices from 84.14 to 102.16 and the grand average hydrophobicity from –0.187 to 0.185 (Supplementary Table 3).

The *GmMlo* gene size ranged from merely 3,485 bp (*GmMlo27*) to 18,992 bp (*GmMlo28*), and 28 of the *GmMlo* genes were over 5,000-bp long (Table 1). The organization of *GmMlo* genes as predicted by GSDS tool is depicted in Supplementary Fig 1. The number of exons varied from 8 (*GmMlo36*) to 18 (*GmMlo34*); most (24) of the *GmMlo* genes had 15 exons, which seems to be a common feature of *Mlo* gene family (Devoto et al. 2003; Singh et al. 2012). Seven *GmMlo* members had 13 exons; four had 14 exons, while 18, 16, 10 and 8 exons were detected in one gene each. The number of amino acids in GmMlo protein predicted using Fgenesh and GENSCAN ranged from merely 288 in *GmMlo36* to 676 in *GmMlo34*, while the size of Mlo domain predicted through Pfam, ranged from only 256 in *GmMlo36* to 619 in *GmMlo34* amino acid residues (Table 1).

Sequence alignment and phylogenetic analysis

The multiple sequence alignment visualized through CLC Sequence Viewer and ClustalW indicated the presence of seven TM regions (Fig. 2). Only the TM3 domain was present in all the GmMlo members, while the other six TM domains were missing/disrupted in one (TM4), two (TM2, TM5), three (TM1, TM7) or four (TM6) of the 39 GmMlos. Comparative phylogenetic trees were constructed

Table 1 List of identified GmMlo members in *Glycine max*

Accession no. GmMlo protein	Gene name	Chr. no.	Gene length (bp)	Amino acid length	Exons	Introns	Number of tm	Mlo domain location	Mlo domain length
XP_003517124.1	<i>GmMlo 1</i>	1	6,218	598	15	14	10	32–520	488
XP_003519926.1	<i>GmMlo 2</i>	2	6,125	573	15	14	7	7–498	491
XP_003520359.1	<i>GmMlo 3</i>	3	6,955	517	13	12	9	6–497	491
XP_003522288.1	<i>GmMlo 4</i>	4	5,518	506	13	12	7	6–495	489
XP_003523525.1	<i>GmMlo 5</i>	4	10,149	595	15	14	7	8–506	498
XP_003523524.1	<i>GmMlo 6</i>	4	4,303	524	15	14	7	55–543	488
XP_003527594.1	<i>GmMlo 7</i>	6	5,044	488	13	12	7	8–479	471
XP_003527634.1	<i>GmMlo 8</i>	6	3,915	541	14	13	7	7–491	484
XP_003532996.1	<i>GmMlo 9</i>	8	4,107	556	15	14	7	6–502	496
XP_003536359.1	<i>GmMlo 10</i>	10	5,191	563	15	14	6	4–473	469
XP_003537693.1	<i>GmMlo 11</i>	11	5,592	600	15	14	10	32–522	490
XP_003538050.1	<i>GmMlo 12</i>	11	8,418	576	15	14	8	5–505	500
XP_003540745.1	<i>GmMlo 13</i>	12	7,842	577	15	14	8	5–502	497
XP_003539407.1	<i>GmMlo 14</i>	12	7,460	582	15	14	7	1–493	492
XP_003543248.1	<i>GmMlo 15</i>	13	10,560	545	15	14	7	6–494	488
XP_003546228.1	<i>GmMlo 16</i>	15	4,859	497	15	14	7	4–482	478
XP_003548777.1	<i>GmMlo 17</i>	16	6,118	583	16	15	10	31–514	483
XP_003548912.1	<i>GmMlo 18</i>	16	5,630	571	15	14	7	6–500	494
XP_003548913.1	<i>GmMlo 19</i>	16	6,445	576	15	14	7	7–498	491
XP_003555433.1	<i>GmMlo 20</i>	20	4,839	559	15	14	6	7–476	469
XP_003516545.1	<i>GmMlo 21</i>	1	4,968	535	15	14	7	6–492	486
XP_003518148.1	<i>GmMlo 22</i>	2	5,662	328	10	9	5	31–327	296
XP_003519925.1	<i>GmMlo 23</i>	2	3,883	430	13	12	5	2–359	357
XP_003520647.1	<i>GmMlo 24</i>	3	7,377	496	15	14	9	6–470	464
XP_003526251.1	<i>GmMlo 25</i>	6	6,609	573	14	13	9	4–487	483
XP_003526073.1	<i>GmMlo 26</i>	6	5,799	513	15	14	7	6–495	489
XP_003527633.1	<i>GmMlo 27</i>	6	3,485	527	15	14	7	9–515	506
XP_003534232.1	<i>GmMlo 28</i>	9	18,992	577	15	14	10	31–514	483
XP_003538881.1	<i>GmMlo 29</i>	11	4,878	530	15	14	7	6–492	486
XP_003539592.1	<i>GmMlo 30</i>	12	6,712	542	15	14	8	6–487	481
XP_003539406.1	<i>GmMlo 31</i>	12	7,860	575	15	14	7	4–500	496
XP_003539400.1	<i>GmMlo 32</i>	12	3,522	429	13	12	5	9–382	373
XP_003540949.1	<i>GmMlo 33</i>	12	6,719	573	14	13	9	4–487	483
XP_003542000.1	<i>GmMlo 34</i>	13	9,677	676	18	17	8	16–635	619
XP_003541720.1	<i>GmMlo 35</i>	13	5,239	496	13	12	6	4–455	451
XP_003546663.1	<i>GmMlo 36</i>	15	5,061	288	8	7	5	32–288	256
XP_003547159.1	<i>GmMlo 37</i>	15	3,803	426	14	13	4	4–426	422
XP_003547751.1	<i>GmMlo 38</i>	16	6,361	515	13	12	9	6–505	499
XP_003553544.1	<i>GmMlo 39</i>	19	6,368	496	15	14	7	6–470	464

The members have been named from 1 to 20 as named by Shen et al. (2012). The additional 19 GmMlo members are named according to their series of location on chromosomes

chr. Chromosome

separately for the *Mlo* genes from *Arabidopsis* (Fig. 3a) and soybean (Fig. 3b); in addition, a combined tree for both *Arabidopsis* and soybean *Mlo* genes was also constructed (Fig. 3c). The MLE method generated dendrograms similar to those produced by the Neighbor-Joining (NJ) method

(phylogenetic trees not shown), and there were only minor variations in the clustering pattern. This variation may be expected due to the differences between the bases of grouping used by the two methods: NJ relies on distance matrix for grouping, while MLE is an evolutionary model-based

method. Therefore, the MLE trees were preferred for inferring evolutionary relationships among the *Mlo* genes (Fig. 3a–c). The phylogenetic trees for *AtMlos* (Fig. 3a), *GmMlos* (Fig. 3b), and the combined *AtMlos* and *GmMlos* (Fig. 3c) had, in each case, three major clusters (clusters I, II and III). The cluster I was the largest and it was subdivided into two sub-clusters (Ia and Ib) so that each tree had a total of four sub-clusters. In the case of *A. thaliana*, cluster Ia had five genes (*AtMlo9*, 10, 5, 7, 8), cluster Ib had four genes (*AtMlo3*, 12, 2, 6), while clusters II and III had three genes each (*AtMlo1*, 15, 13, and *AtMlo4*, 11, 14, respectively). Similarly, the cluster Ia of *G. max* had 17 genes, cluster Ib had 10 genes, while clusters II and III had five and seven genes, respectively. Cluster Ia from the combined phylogenetic tree comprised the cluster Ib of *AtMlos* and the cluster Ia plus two members (*GmMlo6*, 8) from the cluster Ib of *GmMlos*. The cluster Ib of the combined analysis consisted of the cluster Ia of *AtMlos* and

Fig. 2 Multiple sequence alignment of the predicted 39 *GmMlo* proteins. The alignment generated through ClustalW was fed to CLC Sequence Viewer for visualization. The red line block indicates the seven transmembrane regions in *GmMlo* proteins. The last block shows the region of calmodulin-binding domain (CaMBD), an essential characteristic of *MLO* gene family. Note that *GmMlo* 22 and 36 lack TM 5, 6 and 7 along with CaMBD. The CaMBD region shows strong conservation of lysine-rich domain (color figure online)

the remaining eight members of the cluster Ib of *GmMlos*. But the membership of clusters II and III of the combined tree was purely a combination of the members belonging to the respective clusters of *AtMlos* and *GmMlos*. Thus, the combined phylogenetic analysis of *AtMlos* and *GmMlos* produced only a small change in the clustering pattern: the genes *GmMlo6* and *GmMlo8* showed closer relationship with the members of cluster Ia in the combined analysis, while they were placed in cluster Ib when the *GmMlos* were analyzed separately.

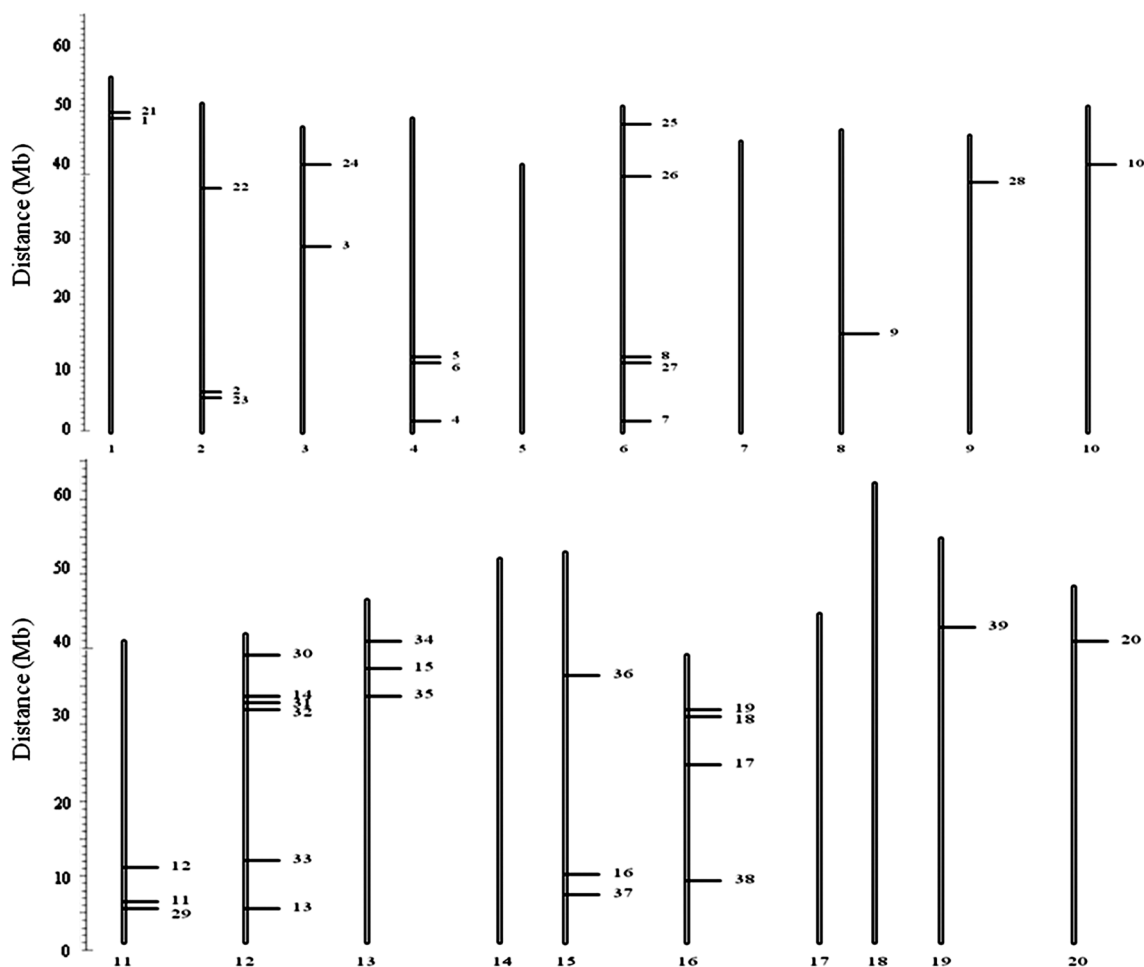
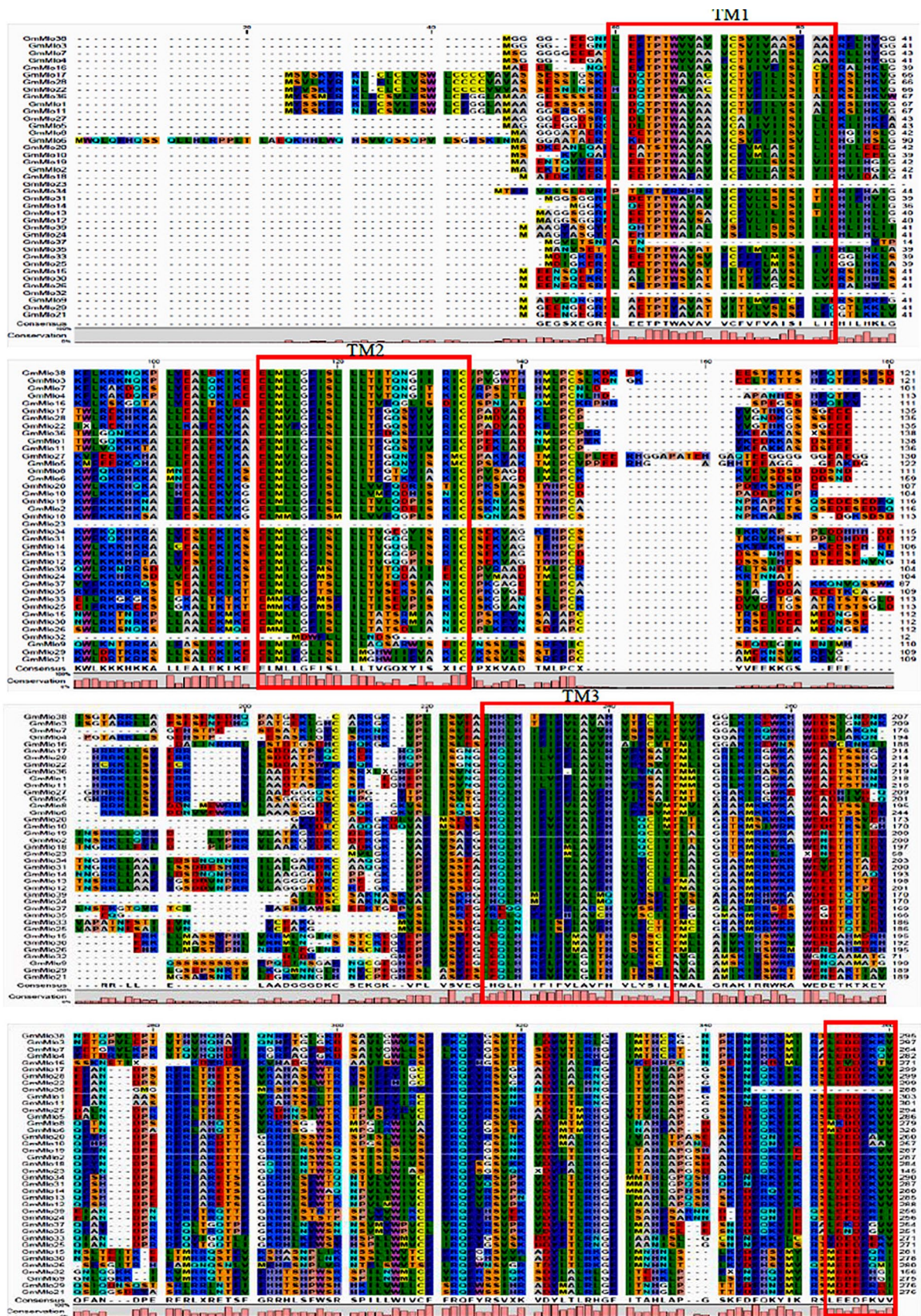


Fig. 1 Physical map of the 39 *GmMlo* members on 15 chromosomes of *G. max*. The size of each chromosome approximately corresponds to the total size of the concerned chromosome. The horizontal line

on each chromosome represents the location of the *GmMlo* gene indicated by a number, corresponding to the gene serial number. Chromosome 5, 7, 14, 17 and 18 do not carry any *GmMlo* gene



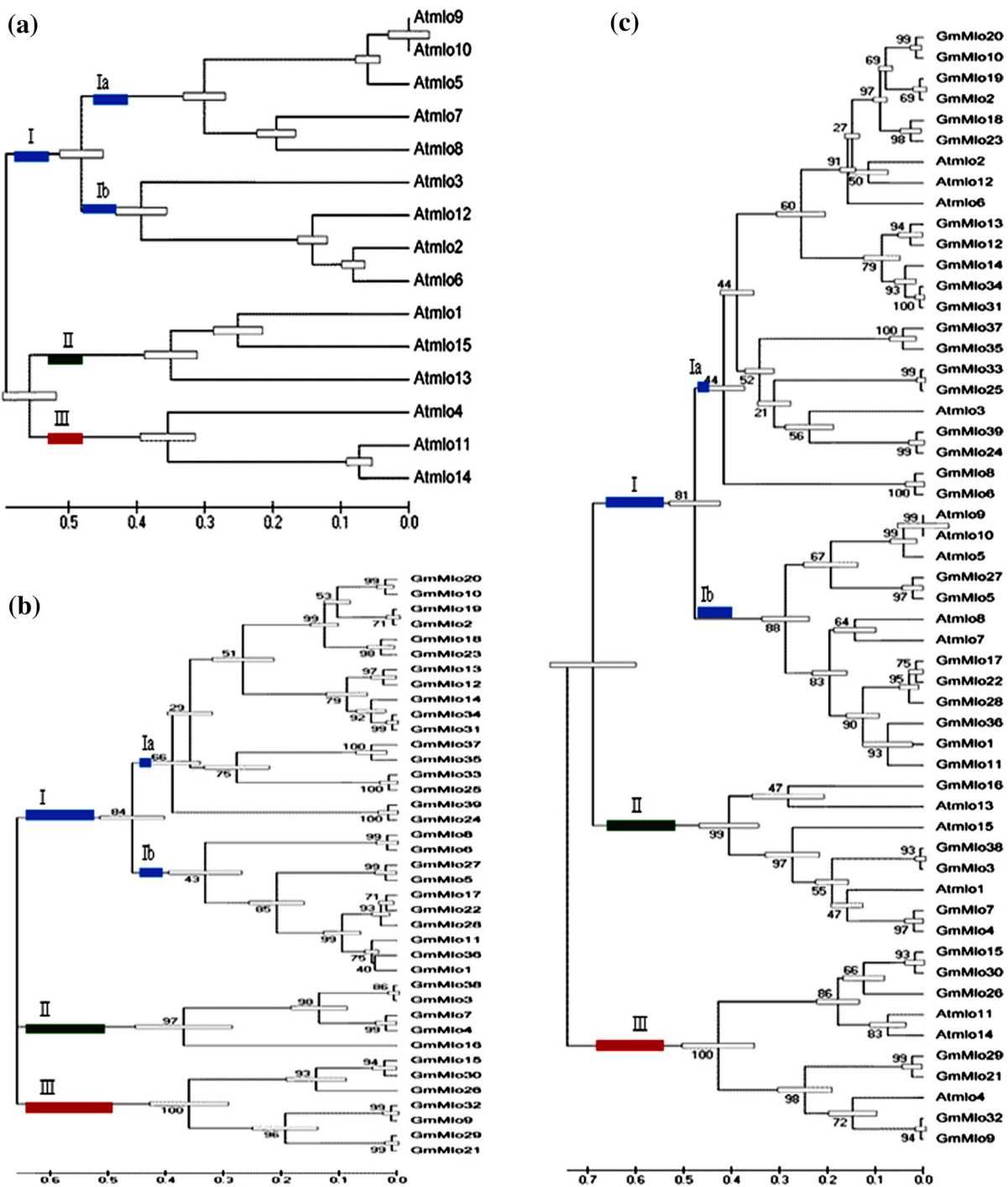


Fig. 3 The AtMlo and GmMlo proteins phylogenetic classification generated using maximum likelihood estimation (MLE) method with the ancestral node at 1,000 replicates. The *Arabidopsis* Mlos are abbreviated with ‘At’ prefix. GmMlo stands for *Glycine max* Mlo. **a** The AtMlo members with three major clusters. **b** The GmMlo members with three major clusters. **c** The comparative AtMlo and GmMlo phylogenetic classification. All the classifications showed formation

of three major clusters (marked with *bold colored lines*, cluster I was divided into two sub-clusters shown in similar *color* as of main cluster). The cluster I shows higher degree of evolutionary expansion with respect to the other clusters in all the linearized trees. The GmMlo members show duplication events that might have occurred after segmental deletion of ancestral gene, as also evident from Geneious pro visualization (color figure online)

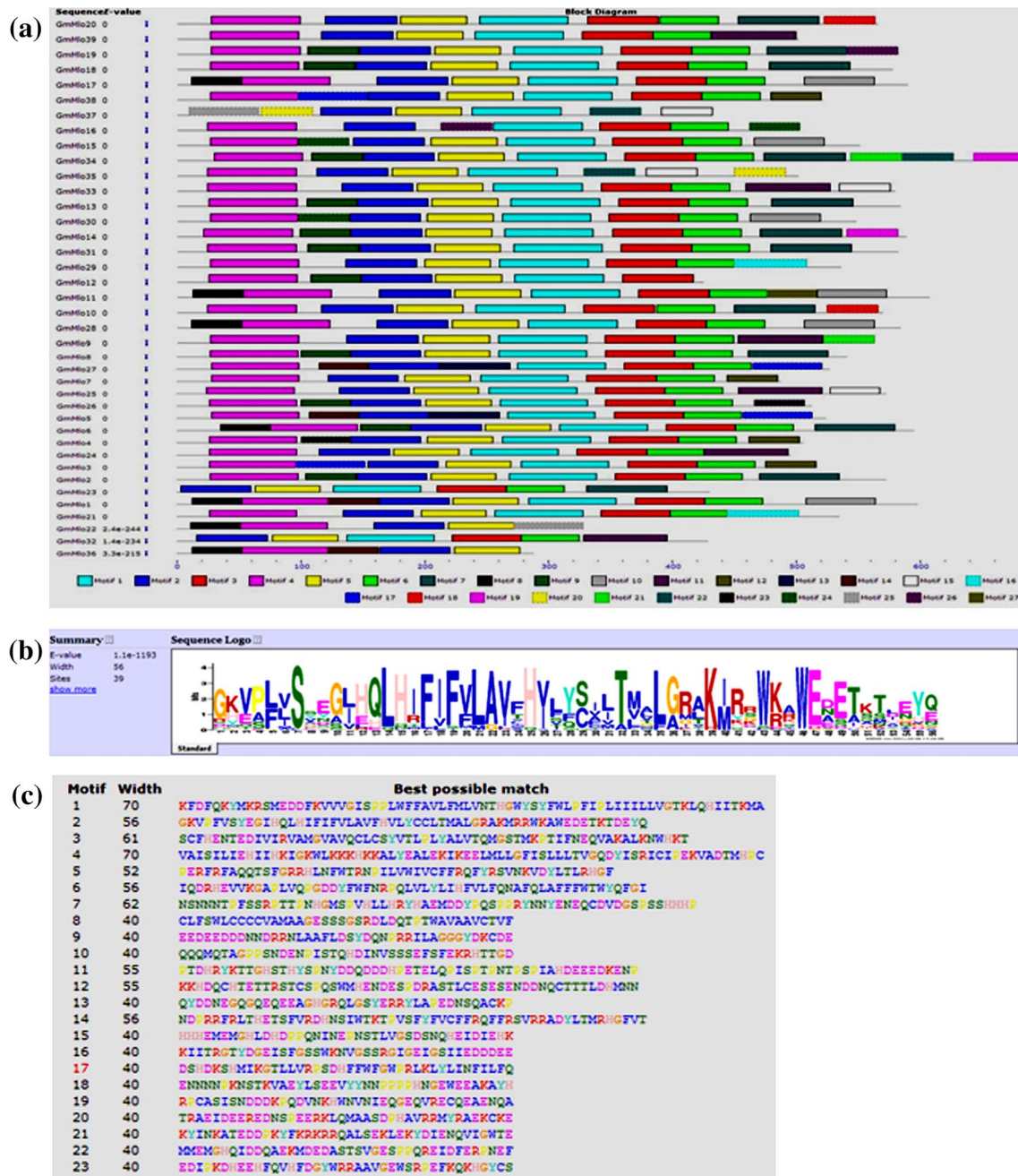


Fig. 4 **a** Combined block diagram of conserved motifs of the predicted 39 GmMlo proteins. **b** The motif 2 was found to be evenly conserved in all the GmMlo proteins. **c** The conserved motif sequences generated using MEME

Discussion

Features of the *Mlo* gene family

The *Mlo* gene family is one of the largest seven TM domain protein-encoding gene families in plants. The *Mlo* genes are known to participate in both constitutive as well as tissue-specific and development-related gene expression. The Mlo proteins function primarily in defense response

to powdery mildew pathogen and programmed cell-death pathways (Peterhänsel et al. 1997), but they are also implicated in various other functions, including responses to abiotic stresses, growth regulators and wounding. In view of the above, this study was initiated to identify and characterize the *Mlo* genes present in soybean genome. We used the 15 AtMlo proteins as query sequences for homology search of the soybean genome sequence available in the NCBI database and identified 39 members of the *Mlo* gene family.

The identified GmMlo proteins showed structural features and cellular localization similar to those of AtMlos, and the *GmMlo* genes exhibited exon–intron boundaries similar to those of *AtMlo* genes. Further, the GmMlo proteins had physico-chemical properties like molecular weight, theoretical pI and amino acid richness more or less comparable to those of AtMlos. Most of the GmMlo proteins were leucine-rich, which is in agreement with the findings for SbMlo (*Sorghum bicolor* Mlo) proteins (Singh et al. 2012). But the proteins GmMlo9, 25, 29, 32 and 33 were serine-rich in the place of being leucine-rich.

In this study, 19 *GmMlo* genes were identified in addition to the 20 genes reported by Shen et al. (2012) based on in silico analysis of soybean genome sequence in the Phytozome database (<http://www.phytozome.net/soybean.php>). We searched the soybean genome sequence available in the Phytozome database (as on 30/10/2013) using the search method of Shen et al. (2012), and identified a total of 35 *GmMlo* members. Similar result was obtained when the Phytozome database was subjected to homology search using AtMlo8, the longest (593 amino acid residues) AtMlo protein, as query sequence. But when AtMlo3 having 556 amino acid residues was used as query sequence for BLAST search, 39 GmMlo members were detected, although four of the members had only partial gene sequence (incomplete protein with as few as 80 amino acid residues). Further, the search of soybean genome sequence in the NCBI database using the method of Shen et al. (2012) identified 39 *GmMlo* members. Thus, the reason for the discovery of 19 additional *GmMlo* members in the present study seems to be the much larger and more complete genome sequence of *G. max* available in the NCBI database (1,115 Mb organized in 20 chromosomes) as compared to that in the Phytozome database (975 Mb distributed in 20 chromosomes).

The 39 *GmMlo* genes covered 0.0897 % of the soybean genome, and were distributed on 15 of the 20 *G. max* chromosomes. Most of the *GmMlo* members occurred as singletons, and only some of them were present in groups of two or three; the members in most of the gene-groups were separated by 20 kb or less. The physical map suggested similarities between pairs of some of the gene groups: for example, the distance between the members of gene-pairs *GmMlo1-GmMlo21* and *GmMlo11-GmMlo29* was virtually identical (0.7 Mb). This suggested that each pair of such *GmMlo* gene groups might have originated by duplication of a single parental genomic region coupled with translocation of one gene group to a different chromosome. Therefore, the synteny between four pairs of such genomic regions (between the regions having gene-groups *GmMlo1-GmMlo21* and *GmMlo11-GmMlo29*, groups *GmMlo2-GmMlo23* and *GmMlo18-GmMlo19*, groups *GmMlo30-GmMlo31* and *GmMlo15-GmMlo 34*, and groups *GmMlo4-GmMlo5-GmMlo6* and

GmMlo7-GmMlo8-GmMlo27) was investigated using the MGCAT (Fig. 5). Both the genic as well as intergenic regions in three (chromosome 1 with 11, 4 with 6, and 12 with 13) of the four comparisons showed extensive synteny at the DNA sequence level. But in the case of chromosomes 2 and 16, a large (~14 kb) insertion/deletion was detected. These findings tend to support the suggestion that these four pairs of genomic regions originated from duplication of four parental genomic segments. *G. max* is a paleopolyploid species that has undergone two duplication events about 59 and 13 million years ago, each followed by gene diversification and loss, and extensive chromosomal rearrangements (Schmutz et al. 2010). The above *GmMlo* gene duplication events might have occurred about 13 million years ago during the second genome duplication event since the concerned intergenic sequences have not greatly diverged. Schmutz et al. (2010) detected homologous genomic regions between all the pairs of chromosomes investigated in this study. Shen et al. (2012) proposed that segmental duplication seems to be the chief method of increase in the number of *Mlo* genes, and findings from the present study are consistent with this suggestion.

Mlo proteins are reported to form 6 or 7 hydrophobic domains that have the potential to form TM helices; this feature is shared by Mlo proteins of all the plant species investigated so far. The N-terminus of TM domains has extracellular location, while their C-termini are exposed in the cytoplasm. The amino acid sequence variation, topology, and localization in plasma membrane of Mlo proteins are reminiscent of the seven TM domain G-protein-coupled receptors of metazoa (Devoto et al. 1999). The GmMlo proteins contained 4–10 TM domains: 8 proteins had less than seven TM domains, while 13 proteins had more than 7 TM domains, indicating that seven TM domains are not essential for the function being performed by these genes. Multiple sequence alignment of the GmMlo sequences showed that all the TM domains, except TM3, were disrupted in one or more of the proteins. But the TM domains 1, 2, 4, 5, 6 and 7 were disrupted/absent in one or more of the five GmMlo proteins that were less than 450 amino acids in length, i.e., presumably the truncated versions of Mlo proteins. Only TM6 was disrupted in one GmMlo protein that was 496 amino acids long. Thus all the TM domains, except for TM6, seem to be conserved in the GmMlo proteins, and appear to be needed for their function. The basic biochemical function of the Mlo proteins is not well understood. The disease susceptible response mediated by the wild-type Mlo protein does not implicate the heterotrimeric G-protein signaling. But it is possible that these proteins function as G-protein-coupled receptors in processes other than pathogen response. Alternatively, Mlo proteins may function as cell surface receptors in a signaling cascade that does not involve the G-protein complex (Lorek et al. 2010).

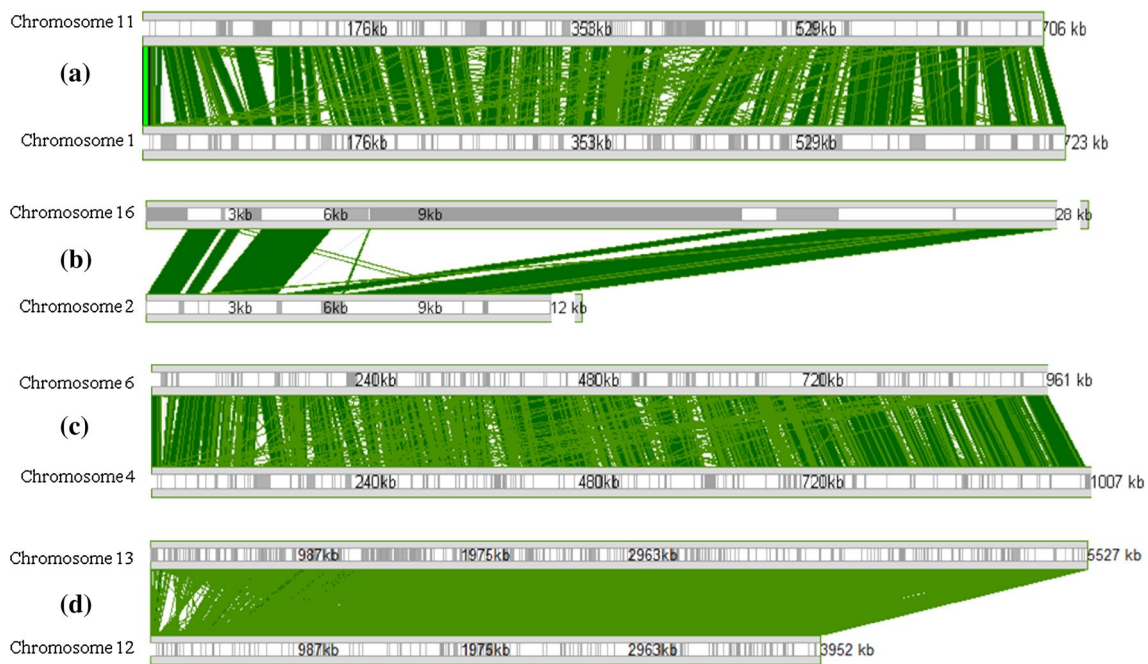


Fig. 5 Synteny between GmMlos containing genomic regions of pairs of *G.max* chromosomes analyzed by MGCAT. **a** Chromosome 1 (GmMlo 1, 21) and chromosome 11 (GmMlo 11, 29). **b** Chromosome 2 (GmMlo 2, 23) and chromosome 16 (GmMlo 18, 19). **c** Chromosome 6 (GmMlo 7, 8, 27) and chromosome 4 (GmMlo 4, 5, 6). **d**

Chromosome 13 (GmMlo 15, 34) and chromosome 12 (GmMlo 30, 31). The syntenic gene sets show collinear gene order between pairs of chromosomes. The above relationship indicates that the syntenic arrangements of respective genes on different chromosomes have resulted due to common ancestry instead of random shuffling

The CaMBD was completely missing in two of the truncated proteins (GmMlo22, 36), but was present in the remaining 37 proteins. Shen et al. (2012) had reported that *GmMlo6* did not have a detectable CaMBD, and suggested that it may be either a special type of *Mlo* gene or a pseudogene. In this study, *GmMlo6* had a CaMBD similar to those of the other members of the family. Further, the protein GmMlo6 identified from the currently available soybean genome sequence in the Phytozome database also had a regular CaMBD. Thus, Shen et al. (2012) failed to detect a CaMBD in GmMlo6 most likely due to incomplete sequence of the gene in the Phytozome database at the time of their search. It was observed that the C-terminal region of CaMBD was much less conserved than its N-terminal portion, indicating that the latter might be necessary for Mlo protein function, especially in disease response. CaMBD is involved in binding of the Mlo protein with the calmodulin protein, which seems to be necessary for Mlo function in powdery mildew susceptibility (Kim et al. 2002a, b; Lecourieux et al. 2006). When Mlo proteins are unable to bind calmodulin, a partially resistant response to powdery mildew is observed (Kim et al. 2002a, b). The *GmMlo22* and *GmMlo36* genes did not have a CaM-binding domain. Therefore, these genes would not be able to participate in the host response to powdery mildew and in regulated cell death (Kim et al. 2002a, b; Bai et al. 2008).

However, the gene *GmMlo36* does contain the AT-rich element in its promoter region for fungal elicitor response. It seems that *GmMlo36* was initially involved in response to biotic stresses due to fungal pathogens, but the deletion of CaMBD would have now impaired this function.

The putative promoter regions of the *GmMlo* genes had a variety of response elements, including seed-specific, nodule site, circadian rhythm, etc. suggesting that they are involved in several developmental and stress response pathways. The *Mlo* genes are reported to participate in programmed cell death, and in response to powdery mildew pathogen (Kim et al. 2002b; Bai et al. 2008), wounding, auxins (Piffanelli et al. 2002; Feechan et al. 2009) and abiotic stresses (Shen et al. 2012). In general, the paralogous members of the family, as indicated by their clustering pattern (Fig. 4b), did not show similarities in their 5' cis-acting response elements. For example, the paralogs *GmMlo17*, 22 and 28 did not share response elements (Supplementary Table 2). Thus the paralogous members of the *GmMlo* gene family share the 'signature motif' of the family, but during the course of evolution their functions seem to have diverged.

Motif and comparative phylogenetic study

The use of Geneious pro offered an easy way to correlate phylogenetic classification with the conserved motif study.

Comparative phylogenetic analysis facilitated clustering of the 39 GmMlo members into four sub-clusters, the tree topology pattern being similar to that of AtMlo members. It appears that the AtMlo and GmMlo members diverged along similar evolutionary lines and their sequence homology and sequence conservation suggest that they might have similar functions. Clustering pattern of the 20 GmMlos studied by Shen et al. (2012) along with the Mlo family members of eight other plant species, including *A. thaliana*, was essentially comparable with the grouping pattern of GmMlos obtained in this study (Fig. 4b). Further, the tree topology of GmMlos (Fig. 4b) suggests that several rounds of duplication events, leading to the present 39 GmMlos, have occurred at different times during the evolution of *G. max*. It appears that there were four different events of gene duplication in the clusters II and III, while seven and eight duplication events seem to have occurred in sub-clusters Ib and Ia, respectively.

Different *GmMlos* seem to have evolved by different mechanisms. For example, GmMlo17 and 22 proteins are very closely related, but GmMlo22 is truncated, while GmMlo17 is a full-length protein. It seems that these two genes originated from a recent duplication event, followed by a large deletion in GmMlo22. Similarly, *GmMlo35* and *GmMlo37* appear to have originated due to a duplication event, followed by deletion in *GmMlo37*. The presence of additional motifs at the C-terminus of GmMlo34 indicates an increase in the sequence of this gene due to duplication of the concerned motif sequences. In this study, both gain and loss of different motifs were observed in the different *GmMlo* members.

The present in silico study attempts to identify and characterize members of the *Mlo* gene family in the *G. max* genome. The Mlo proteins function as modulators of infection by the powdery mildew pathogen, which is reported to attack a large number of crop species. Homology search using AtMlos as query sequences identified 39 *GmMlo* genes located in 15 of the 20 *G. max* chromosomes. In addition, the analysis revealed that these proteins participate in several other functions, including response to abiotic stresses and developmental pathways. The *GmMlo* members seem to be involved in various cellular processes, exhibit tissue- and development-specific expression and many of them are likely to be involved in responses to fungal pathogens. The findings from this study may be helpful in developing strategies for breeding powdery mildew resistance in soybean, and functional validation and isolation of the individual *GmMlo* genes.

Acknowledgments This work was supported by the UGC research grant. We would also like to thank Centre of Bioinformatics; DBT funded SUB-DIC Centre at School of Biotechnology, Banaras Hindu University, Varanasi for providing us the necessary softwares and technical help.

Ethical standard The experiments performed in the above article comply with the current laws of the country in which the experiment has been done.

Conflict of interest The authors declare that they have no conflict of interest.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Bai Y, Pavan S, Zheng Z, Zappel NF, Reinstädler A, Lotti C, De Giovanni C, Ricciardi L, Lindhout P, Visser R, Theres K, Panstruga R (2008) Naturally occurring broad-spectrum powdery mildew resistance in a Central American tomato accession is caused by loss of mlo function. *Mol Plant Microbe Interact* 21:30–39
- Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Second Int Conf Intell Syst Mol Biol*. AAAI Press, Menlo Park, pp 28–36
- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37:W202–W208
- Buchanan SG, Gay NJ (1996) Structural and functional diversity in the leucine-rich repeat family of proteins. *Prog Biophys Mol Biol* 65:1–44
- Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 268:78–94
- Büschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Frijters A, van Daelen R, van der Lee T, Diergaarde P, Groenendijk J, Topsch S, Vos P, Salamini F, Schulze-Lefert P (1997) The barley MLO gene: a novel control element of plant pathogen resistance. *Cell* 88:695–705
- Chen Z, Hartmann HA, Wu MJ, Friedman EJ, Chen J, Pulley M, Schulze-Lefert P, Panstruga R, Jones AM (2006) Expression analysis of the AtMLO gene family encoding plant-specific seven-transmembrane domain proteins. *Plant Mol Biol* 60:583–597
- Chen L, Shiotani K, Togashi T, Miki D, Aoyama M, Wong HL, Kawasaki T, Shimamoto K (2010) Analysis of the Rac/Rop small GTPase family in rice: expression, subcellular localization and role in disease resistance. *Plant Cell Physiol* 51:585–595
- Devoto A, Piffanelli P, Nilsson I, Wallin E, Panstruga R, Heijne GV, Schulze-Lefert P (1999) Topology, subcellular localization, and sequence diversity of the MLO family in plants. *J Biol Chem* 274:34993–35004
- Devoto A, Hartmann HA, Piffanelli P, Elliott C, Simmons C, Tarmino G, Goh CS, Cohen FE, Emerson BC, Schulze-Lefert P, Panstruga R (2003) Molecular phylogeny and evolution of the plant-specific seven-transmembrane MLO family. *J Mol Evol* 56:77–88
- Drummond A, Ashton B, Cheung M, Heled J, Kearse M, Moir R, Stones-Havas S, Thierer T, Wilson A (2008) Geneious v4.0. Biomatters Ltd. Auckland, New Zealand
- Dunleavy JM (1980) Yield losses in soybeans induced by powdery mildew. *Plant Dis* 64:291–292
- Elliott C, Uller JM, Miklis M, Bhat RA, Schulze-Lefert P, Panstruga R (2005) Conserved extracellular cysteine residues and cytoplasmic loop–loop interplay are required for functionality of the heptahelical MLO protein. *Biochem J* 385:243–254
- Feechan A, Jermakow AM, Dry IB (2009) Grapevine MLO candidates required for powdery mildew pathogenicity? *Plant Signal Behav* 4:522–523
- Forsthoefel NR, Cutler K, Port MD, Yamamoto T, Vernon DM (2005) PIRLs: a novel class of plant intracellular leucine-rich repeat proteins. *Plant Cell Physiol* 46:913–922

- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) The proteomics protocols. In: Walker JM (ed) Humana Press
- Guo AY, Zhu QH, Chen X, Luo JC (2007) GSDS: a gene structure display server. *Yi Chuan* 29:1023–1026
- Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database. *Nucleic Acids Res* 27:297–300
- Hoefle C, Huesmann C, Schultheiss H, Börnke F, Hensel G, Kumlehn J, Hüchelhoven R (2011) A barley ROP GTPase ACTIVATING PROTEIN associates with microtubules and regulates entry of the barley powdery mildew fungus into leaf epidermal cells. *Plant Cell* 23:2422–2439
- Huesmann C, Reiner T, Hoefle C, Preuss J, Jurca ME, Domoki M, Fehér A, Hüchelhoven R (2012) Barley ROP binding kinase1 is involved in microtubule organization and in basal penetration resistance to the barley powdery mildew fungus. *Plant Physiol* 159:311–320
- Jørgensen IH (1992) Discovery, characterization and exploitation of Mlo powdery mildew resistance in barley. *Euphytica* 63:141–152
- Jung EH, Jung HW, Lee SC, Han SW, Heu S, Hwang BK (2004) Identification of a novel pathogen-induced gene encoding a leucine-rich repeat protein expressed in phloem cells of *Capsicum annuum*. *Biochim Biophys Acta* 1676:211–222
- Kędzierski L, Montgomery J, Curtis J, Handman E (2004) Leucine-rich repeats in host–pathogen interactions. *Arch Immunol Ther Exp (Warsz)* 52:104–112
- Kemmerling B, Schwedt A, Rodriguez P, Mazzotta S, Frank M, Qamar SA, Mengiste T, Betsuyaku S, Parker JE, Müssig C, Thomma BP, Albrecht C, de Vries SC, Hirt H, Nürnberger T (2007) The BRI1-associated kinase 1, BAK1, Has a brassinolide-independent role in plant cell-death control. *Curr Biol* 17:1116–1122
- Kim DS, Hwang BK (2012) The pepper MLO gene, CaMLO2, is involved in the susceptibility cell-death response and bacterial and oomycete proliferation. *Plant J* 72:843–855
- Kim MC, Lee SH, Kim JK, Chun HJ, Choi MS, Chung WS, Moon BC, Kang CH, Park CY, Yoo JH, Kang YH, Koo SC, Koo YD, Jung JC, Kim ST, Schulze-Lefert P, Lee SY, Cho MJ (2002a) MLO, a modulator of plant defense and cell death, is a novel calmodulin-binding protein. *J Biol Chem* 277:19304–19314
- Kim MC, Panstruga R, Elliott C, Müller J, Devoto A, Yoon HW, Park HC, Cho MJ, Schulze-Lefert P (2002b) Calmodulin interacts with MLO protein to regulate defense against mildew in barley. *Nature* 416:447–450
- Konishi S, Sasakuma T, Sasanuma T (2010) Identification of novel Mlo family members in wheat and their genetic characterization. *Genes Genet Syst* 85:167–175
- Kumar J, Hüchelhoven R, Beckhove U, Nagarajan S, Kogel KH (2001) A compromised Mlo pathway affects the response of barley to the necrotrophic fungus *Bipolaris sorokiniana* (teleomorph: *Cochliobolus sativus*) and its toxins. *Phytopathology* 91:127–133
- Lecourieux D, Ranjeva R, Pugin A (2006) Calcium in plant defense-signalling pathways. *New Phytol* 171:249–269
- Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acid Res* 30:325–327
- Li J, Chory J (1997) A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell* 90:929–938
- Liu Q, Zhu H (2008) Molecular evolution of the *Mlo* gene family in *Oryza sativa* and their functional divergence. *Gene* 409:1–10
- Lorek J, Panstruga R, Hüchelhoven R (2010) The role of seven-transmembrane domain MLO proteins, heterotrimeric G-proteins, and monomeric RAC/ROPs in plant defense. In: Yalovsky S et al (eds) Integrated G proteins signaling in plants, signaling and communication in plants. Springer, Berlin, pp 197–220
- Nibau C, Wu H, Cheung AY (2006) RAC/ROP GTPases: ‘hubs’ for signal integration and diversification in plants. *Trends Plant Sci* 11:309–315
- Opalski KS, Schultheiss H, Kogel KH, Hüchelhoven R (2005) The receptor-like MLO protein and the RAC/ROP family G-protein RACB modulate actin reorganization in barley attacked by the biotrophic powdery mildew fungus *Blumeria graminis* f. sp. hordei. *Plant J* 41:291–303
- Osakabe Y, Maruyama K, Seki M, Satou M, Shinozaki K, Yamaguchi-Shinozaki K (2005) Leucine-rich repeat receptor-like Kinase1 is a key membrane-bound regulator of Abscisic acid early signaling in *Arabidopsis*. *Plant Cell* 17:1105–1119
- Peterhänsel C, Freialdenhoven A, Kurth J, Kolsch R, Schulze-Lefert P (1997) Interaction analyses of genes required for resistance responses to powdery mildew in barley reveal distinct pathways leading to leaf cell death. *Plant Cell* 9:1397–1409
- Piffanelli P, Zhou F, Casais C, Orme J, Schaffrath U, Collins N, Panstruga R, Schulze-Lefert P (2002) The barley MLO modulator of defense and cell death is responsive to biotic and abiotic stress stimuli. *Plant Physiol* 129:1076–1085
- Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD (2012) The Pfam protein families database. *Nucleic Acids Res* 40:D290–D301
- Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R (2005) InterProScan: protein domains identifier. *Nucleic Acids Res* 33:116–120
- Reddy VS, Ali GS, Reddy ASN (2003) Characterization of a pathogen-induced calmodulin-binding protein: mapping of four Ca²⁺-dependent calmodulin-binding domains. *Plant Mol Biol* 52:143–159
- Reinstädler A, Müller J, Czembor JH, Piffanelli P, Panstruga R (2010) Novel induced MLO mutant alleles in combination with site-directed mutagenesis reveal functionally important domains in the heptahelical barley MLO protein. *BMC Plant Biol* 10:31–43
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the paleopolyploid soybean (*Glycine max*). *Nature* 463:178–183
- Schultheiss H, Dechert C, Kogel K, Hüchelhoven R (2002) A small GTP-binding host protein is required for entry of powdery mildew fungus into epidermal cells of barley. *Plant Physiol* 128:1447–1454
- Shen Q, Zhao J, Du C, Xiang Y, Cao J, Qin X (2012) Genome-scale identification of MLO domain-containing genes in soybean (*Glycine max* L. Merr.). *Genes Genet Syst* 87:89–98
- Singh VK, Singh AK, Chand R, Singh BD (2012) Genome wide analysis of disease resistance MLO gene family in sorghum (*Sorghum bicolor* L. Moench). *J Plant Genom* 2:18–27
- Solovyev V, Kosarev P, Seledsov I, Vorobyev D (2006) Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol* 7(Suppl 1):10.1–10.12
- Thompson JD, Higgins DG, Gibson TJ (1994) ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
- Torii KU (2004) Leucine-rich repeat receptor kinases in plants: structure, function, and signal transduction pathways. *Int Rev Cytol* 234:1–46

- Trabanco N, Pérez-Vega E, Campa A, Rubiales D, Ferreira JJ (2012) Genetic resistance to powdery mildew in common bean. *Euphytica* 186:875–882
- Treangen T, Messeguer X (2006) M-GCAT: interactively and efficiently constructing large-scale multiple genome comparison frameworks in closely related species. *BMC Bioinform* 7:433–448
- Tusnády GE, Simon I (1998) Principle governing amino acid composition of integral membrane proteins: application to topology prediction. *J Mol Biol* 283:489–506
- Tusnády GE, Simon I (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics* 17:849–850
- Wolter M, Hollricher K, Salamini F, Schulze-Lefert P (1993) The mlo resistance alleles to powdery mildew infection in barley trigger a developmentally controlled defense mimic phenotype. *Mol Genet* 239:122–128
- Yu CS, Chen YC, Lu CH, Hwang JK (2006) Prediction of protein sub-cellular localization. *Proteins* 64:643–651