Genome-wide identification and comparison of legume MLO gene family

Nicolas Rispail & Diego Rubiales

MLO proteins are highly conserved proteins with seven trans-membrane domains. Specific MLO genes have been linked to plant disease susceptibility. Others are involved in plant reproduction and in root thigmomorphogenesis. Functions of the remaining MLOs are still unknown. Here we performed a genome-wide survey of the MLO family in eight legume species from different clades of the Papilionoideae sub-family. A total of 118 MLO sequences were identified and characterized. Their deduced protein sequences shared the characteristics of MLO proteins. The total number of MLO genes per legume species varied from 13 to 20 depending on the species. Legume MLOs were evenly distributed over their genomes and tended to localize within syntenic blocks conserved across legume genomes. Phylogenetic analysis indicated that these sequences clustered in seven well-defined clades. Comparison of MLO protein sequences revealed 34 clade-specific motifs in the variable regions of the proteins. Comparative analyses of the MLO family between legume species also uncovered several evolutionary differences between the tropical legume species from the Phaseoloid clades and the other legume species. Altogether, this study provides interesting new features on the evolution of the MLO family. It also provides valuable clues to identify additional MLO genes from non-sequenced species.

Grain and forage legumes are among the most important crops worldwide for both animal and human consumptions. They are also important players of sustainable agriculture. Their capacity to fix atmospheric nitrogen allows them to grow in poor soils without application of nitrogenous fertilizers. As a consequence, they contribute to reduce both fossil energy requirement and greenhouse gas emission. Inclusion of legume crops in rotation impacts positively on subsequent crop production. However, legume yield is constantly threatened by fungal diseases.

Powdery mildew emerged as one of the most widespread and damaging legume diseases. One of the most efficient and durable powdery mildew resistance mechanisms was originally found in barley. Lines carrying homozygous recessive alleles at the Mlo locus showed an efficient penetration resistance to this pathogen. mlo-based resistance is one of the few examples of monogenic traits that confer broad spectrum resistance in the field. Recently, mlo-based resistance has been identified in other crops including pea and tomato. Penetration resistance was also detected in other legumes including, Medicago truncatula and Lathyrus belinensis. Although the genetic base controlling the resistance in these species is not known, their phenotypes are reminiscent to that of mlo-based resistant accessions. It is thus possible that they arose from mutations in one MLO gene. Other species may, thus, contain natural mlo mutants that could be very useful to breed crops for resistance to powdery mildew.

The barley MLO is a seven trans-membrane domain protein that localizes at cell plasma membrane. This gene belongs to a highly conserved family found in both monocots and eudicots. To date, the total number of MLO genes identified varied from 11 to 31 according to the plant species (Table 1). The biological function of most MLO genes remains largely unknown. Phylogenetic analyses classified these genes in 6 to 8 clades. All MLO genes with a function in powdery mildew susceptibility clustered in clades IV and V. A clade V MLO gene, from pepper, CaMLO2, has also been associated with susceptibility to bacterial and oomycete pathogens, and to drought. In addition, the expression of some Lathyrus sativus MLO transcripts were induced shortly after rust infection in resistant genotypes, which might suggest their involvement in rust resistance. Although the exact role of clade V MLOs is still unclear, they might interfere with the plant immune response to stresses. This is similar to the Lr34 resistance gene that protects wheat by controlling the induction of multiple defense pathways.
Apart for the known functions of clade IV and VMLOs, two clade I MLO genes from *A. thaliana* (*AtMLO4* and *AtMLO11*) were found to play an important role in root thigmomorphogenesis\textsuperscript{23,24}. Two clade III genes, *AtMLO7* and *OsMLO12*, were also shown to be required for normal pollen tube perception and pollen hydration, respectively\textsuperscript{25,26}, which suggested a role of clade III MLOs in plant reproduction\textsuperscript{14}. The MLO family may thus play a wider range of functions than initially thought. Isolating and characterizing new MLOs from other plant species is thus a promising approach to get new insights on this highly conserved family.

MLO genes have been intensively studied in some plant species. However, little is known about the MLO members of *Fabaceae*. Here, we performed the genome-wide characterization of the MLO gene family in eight legume species belonging to the major clades of the Papilionoideae sub-family (Genistoid, Dalbergioid, Phaseoloid and Galegoid). This included three species from the Galegoid clade (the temperate legumes, barrel medic, chickpea and pea), one from the Genistoid clade (narrow-leaf lupin), one from the Dalbergioid clade (peanut) and three from the Phaseoloid clade that regroups the tropical legumes (pigeon pea, common bean and mung bean) (Table 2). The newly identified sequences were then compared with previously identified MLOs from other plant species to get insights about the evolution of the MLO family in legumes.

### Results

#### Identification of legume MLOs.

Datamining of the different legume genomes (Table 2) identified from 14 to 23 sequences with homology to *A. thaliana* MLOs (Supplementary Table S1). In most genomes, several hits were predicted to encode for truncated proteins. This included the *M. truncatula* sequences *MtMLO12* and *MtMLO16* (Supplementary Table S1). Most of these truncated versions were located close to retro-transposon-like sequences. Thus, these shorter sequences were considered pseudogenes and they were not analysed further.

The remaining sequences were confirmed as putative full length MLOs. This led to the identification of 14 MLO genes in *M. truncatula*, 13 in *Cicer arietinum*, 15 in *Lupinus angustifolius*, 20 in * Cajanus cajan*, 19 in *Phaseolus vulgaris*, 18 in *Vigna radiata* and 13 in each *Arachis* genome (Table 1 and Supplementary Table S1). Interestingly, the sequences SSV2N, from *A. duranensis*, and MQE1N, from *A. ipaensis*, had no counterpart in the second *Arachis* genome (Supplementary Table S1). The peanut genome may, thus, contain 14 potential MLO members. They have been named *ArMLO1* to *ArMLO14* (Table 1 and Supplementary Table S1). To avoid redundancy, only one sequence for each *Arachis* MLO orthologue was used in the analysis.

Since the pea genome has not been sequenced yet, we used the large transcriptomic resources available to search for potential MLO genes in this species. We identified several pea transcripts showing homology with 11 *MtMLO* sequences. This suggested the presence of, at least, 11 potential MLOs in pea (data not shown). In addition to *PsMLO1*, three full length MLO genes could be reconstructed. These sequences, named *PsMLO2, PsMLO3* and *PsMLO4*, showed high similarity to *MtMLO9, MtMLO11* and *MtMLO13*, respectively (Supplementary Tables S1).

#### Organization and distribution of legume MLOs.

The gene characteristics are summarized in Supplementary Table S1. Large variations in gene size were detected within and between legume species. The longest gene, *ArMLO13*, covered a genomic region of 28.05 kb, although this might be due to assembly errors. The

| Plant species | Common name | Total | Clade 1 | Clade 2 | Clade 3 | Clade 4 | Clade 5 | Clade 6 | Clade 7 | Reference |
|---------------|-------------|-------|---------|---------|---------|---------|---------|---------|---------|-----------|
| *Medicago truncatula* (Galegoid clade) | Barrel medic | 14 | 14 | 0 | 1 | 6 | 1 | 1 |
| *Cicer arietinum* (Galegoid clade) | Chickpea | 13 | 13 | 1 | 0 | 6 | 1 | 1 |
| *Pisum sativum* (Galegoid clade) | Pea | 4 | 4 | 0 | 1 | 1 | 1 |
| *Lupinus angustifolius* (Genistoid clade) | Narrow-leaf lupin | 15 | 15 | 0 | 0 | 1 | 1 |
| *Arachis spp.* (Dalbergioid clade) | Peanut | 14 | 14 | 0 | 1 | 1 | 1 | 1 |
| *Glycine max* (Phaseoloid clade) | Soybean | 31 | 31 | 0 | 2 | 9 | 2 | 2 |
| *Cajanus cajan* (Phaseoloid clade) | Pigeon pea | 20 | 20 | 0 | 1 | 1 | 1 | 1 |
| *Phaseolus vulgaris* (Phaseoloid clade) | Common bean | 19 | 19 | 0 | 1 | 6 | 1 | 1 |
| *Vigna radiata* (Phaseoloid clade) | Mungbean | 18 | 18 | 0 | 1 | 6 | 1 | 1 |

Table 1. MLO family members of legume and non-legume species and their phylogenetic classification.\textsuperscript{4} MLO classification is based on Neighbor-Joining phylogenetic analysis and literature. \textsuperscript{4}Actualized *G. max* MLO sequences after removing partial sequences and those classified by Genbank as obsolete sequences.
mean genomic length of the other legume MLOs varied from 4.41 to 6.62 kb. The length of their coding regions varied from 1.63 to 1.65 kb on average distributed on 12 to 17 exons. Accordingly, the mean protein size varied from 539 to 547 amino acids (Supplementary Table S1).

One to four MLO genes were detected on almost all legume chromosomes indicating an even distribution over legume genomes. In addition, we observed that physically close MLO pairs, in any given species, had orthologous pairs in the corresponding chromosome of other legume species (Supplementary Table 1). For instance, the orthologous sequences of MtMLO5, MtMLO6, and MtMLO9, from M. truncatula chromosome 3, are located on the same chromosome in P. vulgaris, Cicer arietinum, V. radiata, Cajanus cajan and Arachis spp. (Supplementary Table S1). Similar situation was found for the orthologues of MtMLO2 and MtMLO6, located in M. truncatula chromosome 5, and those of MtMLO4 and MtMLO7, from chromosome 2, that were detected, in the same order, on the corresponding chromosomes of Cicer arietinum, P. vulgaris and V. radiata, respectively (Supplementary Table S1). This would suggest that at least some of the MLOs localized within syntenic blocks which are conserved across legume genomes.

Characterization of protein and domain organization. MLO proteins are characterized by the presence of seven trans-membrane (TM) domains and one MLO functional domain. To determine whether the legume MLO genes shared these typical characteristics, their deduced amino acid sequences were subjected to different prediction servers (Supplementary Table S2). Almost all sequences were predicted to contain a single MLO domain covering most of the protein length. The sole exceptions were CaMLO5 and VrMLO9 for which two separated MLO domains were predicted (Supplementary Table S3 and Fig. 1). All potential MLOs were predicted to localize within cell membranes (Supplementary Table S3).

The prediction servers used to estimate the number of TM domains (TMHMM27, Psort28 and InterProScan5)29 implemented different algorithms. This lead to some variations in the total number of TM domains predicted (Supplementary Table S3). Despite these small variations, all sequences, except VrMLO3, were predicted to contain seven TM domains. For 97 sequences, the prediction was supported by two or more servers (Supplementary Table S3 and Fig. 1). The TM domain distribution was largely similar between them and it fitted with the distribution of TM domains of typical MLO proteins (Fig. 1). Several putative MLOs were also predicted to contain a signal peptide at their N terminal region (Supplementary Table S3 and Fig. 1).

In parallel, the legume MLO sequences were subjected to the MEME suite server30 to identify conserved amino acid motifs and to uncover species-specific or legume-specific signatures. This identified 14 amino acid motifs common to most MLO sequences (Table 3). These motifs co-localized with the TM domains, the internal loops 2 and 3 and the calmodulin-binding region (CaMBD) (Supplementary Fig. S1). These motifs were also found in most MLO sequences from Glycine max and from non-legume species including Arabidopsis thaliana, Cucumis sativus, Solanum lycopersicum and Vitis vinifera. In addition, they were largely similar and overlapping with the motifs identified in previous studies18,31,32 (Supplementary Fig. S1).

A previous study by Elliot et al.33 identified 30 invariant amino acid residues within 38 MLO sequences. Twenty two of these residues were also invariable in legume MLOs. The other residues were also highly conserved since they only changed in one or two sequences per legume species (Table 4).

Phylogenetic analysis of legume MLOs. The MLO family was previously subdivided in six to eight clades. To classify the legume MLOs, a Neighbor-Joining (NJ) phylogenetic analysis was performed. To this aim, their deduced protein sequences were aligned with already characterized MLO sequences (Table 1 and Supplementary Fig. S2).

This analysis separated the MLO proteins into seven well-supported clades (Supplementary Fig. S2). The MLO members from clade I further clustered in two well-separated sub-clades (Ia and Ib). At least one MLO protein, from each legume species, was found in clade IV that contains the barley MLO. Several members from each legume species also clustered in clade V with the powdery mildew susceptibility genes of Arabidopsis thaliana. Surprisingly, only sequences from tropical legumes clustered in clade VI with AtMLO. By contrast, the last group (clade VII) was nearly exclusively composed of legume sequences except for the tomato protein SIMLO2 (Supplementary Fig. S2).

To confirm this classification and to analyse further the evolution of the MLO family in legumes, a more detailed phylogenetic analysis was performed using the maximum likelihood (ML) or maximum parsimony (MP) algorithms. The two approaches (ML and MP) retrieved very similar tree topologies, thus only the ML tree is shown (Fig. 2). This approach also grouped the legume MLOs in seven clades with clade I further divided in two

| Species | Common name | Legume clade | Depository (Bioproject) | Version | Web address | Reference |
|---------|-------------|-------------|-------------------------|---------|-------------|-----------|
| Medicago truncatula | Barrel medic | Galegoid | JCVI | v.4 | http://jcvi.org/medicago/index.php | Young et al.54 |
| Cicer arietinum | Chickpea | Galegoid | NCBI (PRJNA199909) | v.1 | ftp://ftp.ncbi.nlm.nih.gov.genomes/Cicerarietinum/ | Varshney et al.55 |
| Lupinus angustifolius | Narrow-leaf lupin | Genistoid | NCBI (PRJNA179231) | Draft | http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=AOCW01#contigs | Yang et al.56 |
| Arachis spp. | Peanut | Dalbergioid | PeanutBase | v.1 | http://www.peanutbase.org | Bertolli et al.57 |
| Cajanus cajan | Pigeonpea | Phaseoloid | GigaDB | v.5 | http://gigadb.org/dataset/10028 | Varshney et al.58 |
| Phaseolus vulgaris | Common bean | Phaseoloid | Phytozone | v.1 | http://phytozone.jgi.doe.gov/pz/portal.html?info탈ias= Org_Pvulgaris | Schmut et al.59 |
| Vigna radiata | Mungbean | Phaseoloid | NCBI (PRJNA243847) | v.1 | http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=JJMO01#contigs | Kang et al.60 |

Table 2. Legume genomic databases used in this study.
well-supported branches (cluster Ia and Ib) (Fig. 2). Clades I, II and III were represented by three to four MLOs per legume species. By contrast, clades IV, VI and VII were only represented by one sequence per species (Fig. 2 and Table 1). As already observed after the NJ phylogenetic analysis, clade VI only contained MLO sequences from tropical legumes including *G. max*, *Cajanus cajan*, *P. vulgaris* and *V. radiata* (Fig. 2 and Supplementary Fig. S2). The ML phylogenetic tree also showed the expansion of clade V MLOs in tropical legumes. In these Phaseoloid species, six clade V genes were detected while the other legume species had only two genes (Fig. 2 and Table 1).

Conservation of MLO members within clades and identification of clade-specific motifs. To determine the presence of clade-specific motifs within legume MLOs, they were classified according to the phylogenetic tree and subjected to MEME (Fig. 3). In parallel, all MLO orthologues were aligned with Clustal W to visually assess their overall conservation and to locate the conserved motifs (Supplementary Figs S3 to S9). The MEME analysis revealed 34 clade-specific motifs (Fig. 3B and Table 5). According to this analysis, each clade can be recognized by the presence of two to six motifs. These motifs mostly localised within the first extracellular loop, the second intracellular loop and the C-terminal region (Fig. 3 and Supplementary Figs S3 to S9). Among the clade-specific motifs, six were specific to clade V. Three of these motifs localised at the C terminal end of the proteins. Among them, domain 25 and 27 covered the three distinctive regions identified in the MLOs associated
Table 3. Conserved motifs common to all sequences as detected by MEME software.

| Motif | Sequence | e-value | No. sequence |
|-------|----------|---------|--------------|
| 1     | RSL[EDA]ETPTWA[VAC][TF][V][F][L][V][L][A][L][V][S] | 6.4e-1350 | 208/212 |
| 2     | [LA][ILV][ER][H][KR][S][H][KR][L][GK][W][KK][K][H][N][K][A][L][Y][F] | 1.6e-147 | 205/212 |
| 3     | E[AS]LEK[IV][K][E][A][E][L][M][G][F][S][L][L][V][F] | 1.6e-1304 | 209/212 |
| 4     | [S][A][KR][IC][V][P][S][E][S][K][V][A][D][N][S][S][T][M][W][L][H][F][P] | 1.4e-121 | 193/212 |
| 5     | P[LF][V][L][S][V][E][G][L][H][E][Q][L][H][F][R][F][V][F][L][A][V][I][T][H][L][Y][F][S][C][V][A][L][T][M][V][L][A][J][G][R][A][R][K][M][A] | 3.0e-12726 | 211/212 |
| 6     | W[K][R][K][A][W][E][A][D][T][E][K][S][T][S][L][L][Y][E][Q][E][P][A][S][Y][N][H][D][P][E][S][R][F][F][L][A][T][K][H][E][Q][T][S][T][F][G][V][R][H][I][L][S][H] | 2.5e-1288 | 127/212 |
| 7     | C[S][P][F][Q][Y][F][G][S][T][K][R][V][J][A][D][Y][T][A][L][R][H][G][F] | 1.2e-12795 | 212/212 |
| 8     | K[F][N][P][H][Q][K][Y][M][K][R][S][L][M][E][D][D][F][K][V][V][G][Y][S][P][P][L][W][F][A][F][S][I][V][A][V][V] | 2.9e-12704 | 210/212 |
| 9     | N[Y][T][H][N][G][W][H][Y][N][S][T][Y][W][L][S][P][A][F][P][L][V][I][L][I][L][V][L][A][G][T][K][Q][E][H][J][V][J][T][A] | 1.9e-12701 | 211/212 |
| 10    | M[A][G][L][E][R][D][I][Q][T][E][D][R][K][H][G][A][E][V][I][K][Q][G][[P][L][I][V][V][Q][E][P][S][G][D] | 1.6e-12756 | 156/212 |
| 11    | P[W][N][R][K][P][Q][R][L][V][L][F][H][L][H][F][L][I][F][L][Q][N][A][F] | 1.7e-12609 | 212/212 |
| 12    | [A][T][F][Y][P][F][L][W][T][S][W][W][W][E][Q][T][P][G][F][D][S][C][P] | 3.8e-12746 | 207/212 |
| 13    | R[L][V][A][I][L][M][G][V][A][I][Q][V][F][L][C][S][F][V][S][L][P][L][A][L][T][V][T][M][G][S][T][I][R][M][K] | 7.8e-12792 | 212/212 |
| 14    | K[A][T][S][I][F][D][N][E][Q][R][V][T][A][L][K][K][N][G][W][H][K][T][A][L][A][V][K][K][A][K][H][K][G][S][S] | 2.2e-12728 | 189/212 |

with powdery mildew susceptibility16,35 (Supplementary Fig. S7). On the other hand, four motifs (motifs 4 to 7) could distinguish between the two sub-types of clade I MLOs. Other motifs distinguished between two sub-types within the MLO members of clades II and VII (Fig. 3B). Interestingly, one of the clade VII sub-types was characterized by the presence of two motifs (motifs 33 and 34) that were only detected in Phaseoloid species (Fig. 3 and Supplementary Fig. S9). Many of these motifs were also found in non-legume species including Vitis vinifera, Cucumis sativus and Solanum lycopersicum. This includes all clade V-specific motifs. Nevertheless, 12 motifs (motifs 5, 10, 11, 14, 15, 16, 19, 20, 21, 29, 30 and 32) were only found in legume sequences (Table 5).

On the other hand, the level of conservation of specific amino acid residues was also compared between clades (Table 4 and Supplementary Table S4). This comparison indicated that five of the 30 residues, previously described as variable33, were modified on a clade specific basis (Table 4). For instance, the barley E135 residue was systematically changed to a D in clade VI sequences. Similarly, the Y243 and F135 were changed, in clade VII sequences, to F and S, respectively. Most of these punctual changes (three out of five punctual changes) were detected in clade VII. In the other clades, these amino acid positions remained invariant, as previously described33 (Table 4).

We also challenged the specificity of the 73 amino acids that discriminated between the powdery mildew susceptibility MLO of monocots and eudicots, according to Appiano et al.18 (Supplementary Table S4). As indicated previously, these sequences belonged to two different clades (clade IV and V), Thus we aimed not to validate these specific changes but also to determine whether they could discriminate between monocot and eudicot sequences or between clades. For this, we aligned all clade IV and V legume sequences with the sequences included in Appiano et al.18 (Supplementary Table S4). Thirty six of these amino acid positions differed between clade IV and V sequences. For instance, the monocot valine and serine residues at position 32 and 145, were found unchanged in all clade IV legume sequences. For instance, the barley E135 residue was systematically changed on a clade specific basis (Table 4). For instance, the barley E135 residue was systematically changed to a D in clade VI sequences. Similarly, the Y243 and F135 were changed, in clade VII sequences, to F and S, respectively. Most of these punctual changes (three out of five punctual changes) were detected in clade VII. In the other clades, these amino acid positions remained invariant, as previously described33 (Table 4).

Discussion

MLO is a large protein family highly conserved across plant kingdom. Apart from the well-documented role of some MLOs in powdery mildew susceptibility, the biological functions of MLOs remain largely unknown14. Besides providing hints on their potential functions, studying the diversification and multiplication of MLOs in a given species may give clues on its genome evolution. Thus we performed a genome-wide characterisation of the MLO family in eight legume species belonging to different clades and ecological habitats.

Mining legume genomic databases allowed the identification and characterization of 118 MLO sequences. The total number of MLO sequences varied from 13 in chickpea to 20 in pigeon pea (Table 1 and Supplementary Table S1). This is broadly similar to the situation found in other eudicot species that demonstrated the presence of 15 MLO genes in A. thaliana16, 17 in grapevine17 and tomato18, 14 in cucumber19, 18 in strawberry18 or 19 in peach20. The highest number of MLOs was identified in soybean with 31 full length MLOs18 (Table 1). The phylogenetic analysis showed, in most cases, a pair of soybean MLO genes clustering together, for any given MLO orthologue (Fig. 2). Thus the large MLO expansion in soybean is likely consequence of its recent genome duplication13. We also detected the presence of shorter truncated sequences with homology to MLO genes. Since most of...
them were close to retrotransposon-like sequences, we concluded that they were inactive pseudogenes. However, shorter MLO-like sequences have been described in many plant species including tomato 31, cucumber 32, soybean15, strawberry17 and apple17. Thus, these sequences might represent a new family of membrane-proteins not considered before.

The MLO genes were widely distributed over the legume genomes. They were found on almost all chromosomes of any given species (Supplementary Table S1). In addition, most legume MLO orthologues were located, within conserved syntenic blocks, in related chromosomes in the different legumes (Supplementary Table S1). The MLO distribution supported the high level of micro- and macro-synteny that exist between legume genomes38,39. It also further illustrated the assumption that most legume genes are likely located in syntenic regions40 as previously demonstrated for most phenylpropanoid genes of soybean and common bean41. This distribution also suggested that they mainly arose from segmental duplication as it was already assumed for rice and several Rosaceae species17,42. Tandem duplication may have also played a minor role in MLO evolution since we detected evidence of a few tandem duplication events such as the gene pairs PvMLO5/PvMLO6 and VrMLO5/VrMLO6 in P. vulgaris and V. radiata, respectively (Supplementary Table S1).

The phylogenetic analysis classified the legume MLOs in seven clades (Figs 2 and 3), which is in accordance with previous studies4,14,37. The largest clades were clades I, II and III that contained two to six MLO genes, in each legume species (Fig. 2 and Supplementary Fig. S2). In our analysis, the clade I was further divided in two well-supported sub-clades. These sub-clades can also be distinguished by the examination of their sequences (Fig. 2 and Supplementary Figs S2 and S3). One MLO per legume species was found in clade IV that was originally thought to be restricted to monocots14. This distribution also suggested that they mainly arose from segmental duplication as it was already assumed for rice and several Rosaceae species17,42. Tandem duplication may have also played a minor role in MLO evolution since we detected evidence of a few tandem duplication events such as the gene pairs PvMLO5/PvMLO6 and VrMLO5/VrMLO6 in P. vulgaris and V. radiata, respectively (Supplementary Table S1).

The phylogenetic analysis classified the legume MLOs in seven clades (Figs 2 and 3), which is in accordance with previous studies4,14,37. The largest clades were clades I, II and III that contained two to six MLO genes, in each legume species (Fig. 2 and Supplementary Fig. S2). In our analysis, the clade I was further divided in two well-supported sub-clades. These sub-clades can also be distinguished by the examination of their sequences (Fig. 2 and Supplementary Figs S2 and S3). One MLO per legume species was found in clade IV that was originally thought to be restricted to monocots14. This distribution also suggested that they mainly arose from segmental duplication as it was already assumed for rice and several Rosaceae species17,42. Tandem duplication may have also played a minor role in MLO evolution since we detected evidence of a few tandem duplication events such as the gene pairs PvMLO5/PvMLO6 and VrMLO5/VrMLO6 in P. vulgaris and V. radiata, respectively (Supplementary Table S1).

The phylogenetic analysis classified the legume MLOs in seven clades (Figs 2 and 3), which is in accordance with previous studies4,14,37. The largest clades were clades I, II and III that contained two to six MLO genes, in each legume species (Fig. 2 and Supplementary Fig. S2). In our analysis, the clade I was further divided in two well-supported sub-clades. These sub-clades can also be distinguished by the examination of their sequences (Fig. 2 and Supplementary Figs S2 and S3). One MLO per legume species was found in clade IV that was originally thought to be restricted to monocots14. This distribution also suggested that they mainly arose from segmental duplication as it was already assumed for rice and several Rosaceae species17,42. Tandem duplication may have also played a minor role in MLO evolution since we detected evidence of a few tandem duplication events such as the gene pairs PvMLO5/PvMLO6 and VrMLO5/VrMLO6 in P. vulgaris and V. radiata, respectively (Supplementary Table S1).

The phylogenetic analysis classified the legume MLOs in seven clades (Figs 2 and 3), which is in accordance with previous studies4,14,37. The largest clades were clades I, II and III that contained two to six MLO genes, in each legume species (Fig. 2 and Supplementary Fig. S2). In our analysis, the clade I was further divided in two well-supported sub-clades. These sub-clades can also be distinguished by the examination of their sequences (Fig. 2 and Supplementary Figs S2 and S3). One MLO per legume species was found in clade IV that was originally thought to be restricted to monocots14. This distribution also suggested that they mainly arose from segmental duplication as it was already assumed for rice and several Rosaceae species17,42. Tandem duplication may have also played a minor role in MLO evolution since we detected evidence of a few tandem duplication events such as the gene pairs PvMLO5/PvMLO6 and VrMLO5/VrMLO6 in P. vulgaris and V. radiata, respectively (Supplementary Table S1).

The phylogenetic analysis classified the legume MLOs in seven clades (Figs 2 and 3), which is in accordance with previous studies4,14,37. The largest clades were clades I, II and III that contained two to six MLO genes, in each legume species (Fig. 2 and Supplementary Fig. S2). In our analysis, the clade I was further divided in two well-supported sub-clades. These sub-clades can also be distinguished by the examination of their sequences (Fig. 2 and Supplementary Figs S2 and S3). One MLO per legume species was found in clade IV that was originally thought to be restricted to monocots14. This distribution also suggested that they mainly arose from segmental duplication as it was already assumed for rice and several Rosaceae species17,42. Tandem duplication may have also played a minor role in MLO evolution since we detected evidence of a few tandem duplication events such as the gene pairs PvMLO5/PvMLO6 and VrMLO5/VrMLO6 in P. vulgaris and V. radiata, respectively (Supplementary Table S1).
specific incorporation in the genome of the Phaseoloid species. Lupin and peanut belongs to the early-diverging clades, Genistoid and Dalbergioid, respectively. Their separation from the other Papilionoidoids clades has been
The evolution of the tropical (Phaseoloids) and temperate legumes (Galegoids) is more recent\textsuperscript{43,44}. It has been estimated to have taken place approximately 52.8 and 50 million years ago\textsuperscript{44}. According to this, it appears more likely that the Phaseoloid species (common bean, pigeon pea, mung bean and soybean) have incorporated this MLO clade during speciation. The phylogenetic study also revealed a seventh clade that was mainly represented by legume MLO. This is in accordance with recent studies that also identified a seventh clade in cucumber\textsuperscript{32} and tomato\textsuperscript{31}. Another recent study on Rosaceae MLO identified two new clades apparently restricted to Rosaceae species (clades VII and VIII)\textsuperscript{17}. However, MLO sequences from soybean, cucumber or tomato were not included in their analysis\textsuperscript{17}. Thus, a more global analysis of MLO sequences, over plant kingdom, would be necessary to determine whether evolution of MLO sequences led to the apparition of genera-specific clades.

Previous studies identified several conserved motifs\textsuperscript{15,16,31,32,33} that we also detected in the legume MLO protein sequences (Table 3 and Supplementary Fig. S1). One of these common motifs, located at the C-terminal

---

**Figure 3. Motif organization of legume MLOs.** The figure shows the predicted topology of a typical MLO protein (a) and the schematic organization of the common and specific motifs for each MLO clade (b). Common and clade-specific motifs are represented by white and colored boxes, respectively. These motifs were identified by scanning the MLO sequences with the MEME suite software\textsuperscript{30} (Supplemental Table S1). Common and clade-specific amino acid motifs are listed in Tables 3 and 4 respectively. Localization of transmembrane domains is shown as dashed horizontal lines.
region, was previously shown to bind to the calcium-sensing protein calmodulin. Here, we confirmed that the calmodulin binding site was conserved in all MLO clades (Supplementary Figs S3 to S9), since it was found within the common conserved motif 14 in all legume sequences (Table 3). In addition to these common motifs, our study identified 34 clade-specific motifs and several clade-specific amino acid residues. These motifs located in the extracellular loops 1 and 3, the intracellular loop 2 and the C terminal region (Table 4, Fig. 3 and Supplementary Table S4). For instance, six clade V-specific motifs were detected (Table 4, Fig. 3 and Supplementary Fig. S7). Two of them, motifs 25 and 27, contained the previously identified consensus clade V sequences. This confirmed the efficacy of the method used. Interestingly, the conserved tetrad E/D FSF was also detected in clade IV and VI MLO sequences (Supplementary Figs S6, S7 and S8). The presence of this motif in clade IV sequences region, was previously shown to bind to the calcium-sensing protein calmodulin. Here, we confirmed that the calmodulin binding site was conserved in all MLO clades (Supplementary Figs S3 to S9), since it was found within the common conserved motif 14 in all legume sequences (Fig. 3). In addition to these common motifs, our study identified 34 clade-specific motifs and several clade-specific amino acid residues. These motifs located in the extracellular loops 1 and 3, the intracellular loop 2 and the C terminal region (Table 3). In addition to these common motifs, our study identified 34 clade-specific motifs and several clade-specific amino acid residues. These motifs located in the extracellular loops 1 and 3, the intracellular loop 2 and the C terminal region (Table 4, Fig. 3 and Supplementary Table S4). For instance, six clade V-specific motifs were detected (Table 4, Fig. 3 and Supplementary Fig. S7). Two of them, motifs 25 and 27, contained the previously identified consensus clade V sequences. This confirmed the efficacy of the method used. Interestingly, the conserved tetrad E/D FSF was also detected in clade IV and VI MLO sequences (Supplementary Figs S6, S7 and S8). The presence of this motif in clade IV sequences may have been expected since it contains the powdery mildew susceptibility genes of monocots. By contrast, its presence in the more divergent clade VI is surprising. It might indicate a common mechanism of action of these three clades. The identification of clade-specific motifs is very useful to isolate MLO orthologues in plant species not yet sequenced.

Beyond finding interesting new features about the MLO gene family, our study also showed diverging features between the tropical legumes (Phaseoloids) and the other legume species. One of the most striking differences was the total number of MLOs found in each type of legumes. Legumes from the Genistoid, Dalbergioid and Galegoid clades were characterized by 13 to 15 genes while tropical legumes contained from 18 to 31 genes (Table 4 and Supplementary Table S1). Almost all additional genes from tropical legumes clustered in clade V. The total number of MLOs in tropical legumes may reflect a greater pathogenic variability and pressure in tropical regions.

Table 5. Clade-specific motifs in MLO sequences as detected by MEME software.

| Clade | Motif | Sequence |
|-------|-------|----------|
| I     | 1     | [DN][SG][LN][SE][QI][IS][TK][IK][ES][LK][TR][ML][RR][QL][ST][TF][VI][FK][HS][HT][SH][WJ][KL][SR][HL][KH][KSP] |
| I     | 2     | F[IM][NE][HLN][PS][KL][T KS] |
| I     | 3     | [TK][LA][LV][EN][A][GE][QR][TC][GP][TP][MM][SK][ELL][HT][KF][LQ][RF][KR][PL][RDEL][WF][KNG][KP][ERD] |
| I     | 4     | TRS[EQI][ED][E][MI][E][ED][NGSE][E][RR][L][MT][A] |
| I     | 5     | [YA][PY][HR][LV][IF][RG][RL][ND][GI][MN][NR][SN][ST] |
| I     | 6     | THDDTSTVLSIEEDDDQ[D][DAT][PE] |
| I     | 7     | [AT][VT][SA][TP][PL][RS][AS][SS][LY][HQ][V][P][CS][SF][S] |
| I     | 8     | G[AL][RK][IR][QE][KH][W][WD][Q][PE][Q][LE][Y][HE][XT] |
| I     | 9     | [RF][VL][KE][P][KT][VT][HN][V][HQ][QH][ADE][FI] |
| I     | 10    | G[QLG][V][PR][KK][KR][AR][SA][AS][PA][Q][E][DE] |
| I     | 11    | [AKN][KR][IK][KR][GL][LK][RS][AG][D][SGN][NQ][SQ][SH][GP][HG][GS][S] |
| I     | 12    | EG[EG][EH][RH][FR][LL][SR][YF][IL][AS][AG][ED][GTA][TG][SG] |
| I     | 13    | DGS[ST][VI][HS][G][PT][LR][HT][K][TG][HT][GR][SR] |
| I     | 14    | Y[DE][Q][D][D][DE][Y][HE][SQ][SD][E] |
| I     | 15    | [PO][T][AT][SN][LI][LV][IV][Y][R][VD][HN][G][DE][Q][Q][AE][Q][EE][EN][HE][H] |
| I     | 16    | [FD][DE][NE][ML][MV][IF][V][W][R][VL][LA] |
| I     | 17    | A[AS][SG][GD][DC][Y][Q][GK][GP][LS][Q][SG][V] |
| I     | 18    | SGE[TA][T][PSG][TH][NI][H][H][Q][Y][K][P][S] |
| I     | 19    | [H][T][T][D][S][D][V][Y][L][P][R][S][Q][SD] |
| I     | 20    | [TR][N][G][Q][EL][TH][N][DE][PS][VF][Y][K][P][S] |
| I     | 21    | [fi][Li][S][D][TE][G][S][TH][QL][EL][U][Q][T][Q][T][M][HS][Q][P]A] |
| I     | 22    | P[DN][G][Q][EL][TH][N][DE][PS][VF][Y][K][P][S] |
| I     | 23    | LA[T][AG][IK][G][E][D][R][C] |
| I     | 24    | FW[TS][QR][SN][PT][IV][LS][LV][W][W][V]WIV |
| I     | 25    | FHSTT[EA][D][V][V][V][I][V][R][I][L][] |
| I     | 26    | SPP[SSR][STPT][T][H][G][MS][P][Y][H][ALL] |
| I     | 27    | [AP][GR][RS][RH][ED][SD][AF][Q][PS][NR][SN][ED][NE][Q][WD] |
| I     | 28    | P[T][F][SR][ST][Q][E][H][IL][HI][KN][HR][IN][IV][SA][SL][SK][ED][PS][VF][ED][KR][RG][HP][TH] |
| I     | 29    | [QR][NE][LO][Q][E][L][Q][V][K][S][RS][IF][HR] |
| I     | 30    | MSS[WS][Q][GS][KN][AP][ED][F][ST][L][CR][E][SN][I][ANT][RSS][D][E][E][V][E][E][E][E][E][L][H][MP][S][V][V][K][R][Q][W][H][V][L][E][V][R][ER] |
| I     | 31    | [N][P][K][R][TR][G][T][I][Y][D][E][H][G][F][S][Y][W][W][K][G][N][S] |
| I     | 32    | S[SK][GR][E][G][V][S][I][T][A][I][EE][D][E][D] |
| I     | 33    | AT[RI][HS][TS][EG][L][E][D][V][I][A][PH][AT][N][E][Q][S][T][A][E][N][IV][I][E][V][Y][F] |
| I     | 34    | NN[ST][K][H][S][D][L][H][E][K][G][C][DNS][AV][RG][ST][VM][DS][V][TH][P][D][N][V][VA][LV][T][S][N][NP][P][S][F][PH] |

Table 5. Clade-specific motifs in MLO sequences as detected by MEME software.
family evolved differently depending on the legume clade considered. The tropical legume diverged after the separation from lupin and peanut ancestors but before that of the temperate legumes. They are the only ones to have incorporated the clade VI MLO in its genome and to have followed a significant expansion of this family (Fig. 2 and Supplementary Fig. 2). Our data also indicate that the Genistoid, Dalbergoid and Galegoid, have evolved in parallel. These assumptions were also supported by the detection of two specific motifs (motifs 33 and 34) only found in clade VII MLOS of tropical legumes.

Altogether, our study characterized 118 MLO sequences from eight different legume species with different habitats and agronomic characteristics. This comparative analysis revealed interesting new phylogenetic features that may be the base to further determine the function of this gene family. We also detected several differences between tropical and the other legume species that might reflect different evolutionary pressures. In addition, we identified from three to seven genes in clades IV and V that contains the genes associated with powdery mildew susceptibility. These new sequences are very valuable to identify new gene variants to confer powdery mildew resistance in these species and to identify new susceptibility genes in additional legume species.

Methods
Identification, annotation and validation of legume MLOS. M. truncatula MLO sequences were identified by mining the JCVI M. truncatula genomic project v4.0 database through BLAST searches with the 15 Arabidopsis thaliana MLO sequences as templates. All potential MLO sequences from the other legume species (Table 1) were retrieved by BLAST45 using the M. truncatula MLO CDS and protein sequences. In all cases, the lowest limit of significance (e-value) for any potential hits was set at 1e−20. All potential MLO sequences were systematically validated by reciprocal BLAST on the M. truncatula JCVI Mt4.0 (http://jcvi.org/medicago/index.php) and A. thaliana TAIRv10 databases (http://www.arabidopsis.org).

Upon validation, the genomic sequence of each potential MLO was examined to reconstruct the full length CDS and correct potential annotation errors. Each genomic sequence was then aligned to its corresponding transcripts by BLAST against its respective transcriptomic (TSA) and EST databases that are stored at the NCBI website (http://blast.ncbi.nlm.nih.gov/Blast.cgi). In parallel, validated MLO sequences, from unannotated legume genomes (Lupinus angustifolius and Vigna radiata), were analyzed with Fgenesh46 using the “Medicago legume gene” model (Supplementary Table S2). Manual correction of the annotation was also performed, if necessary, to improve sequence quality. SeqBuilder v12.0 (DNAStar, Madison, WI) was used to draw and correct the resulting exon-intron organization of each sequence. Supplementary Figs S10 and S11 show the CDS and deduced protein sequences of the legume MLOS, respectively.

Conservation and phylogenetic analyses. Global pair-wise analysis was performed on the deduced protein sequences to determine their level of conservation to their closest homologue in M. truncatula and A. thaliana with Geneious R8 (http://www.geneious.com)47. Multiple protein sequence alignments were performed with ClustalW48. The alignments were manually corrected before phylogenetic reconstruction.

To assign each potential MLO to its clade, all identified MLO protein sequences were aligned with the MLO sequences from soybean and several non-legume species (Table 1). This alignment was used to calculate a distance matrix with the NJ algorithm. This analysis was performed with 1,000 bootstrap replicates. Prior to ML analysis, all gaps and divergent regions were removed from the protein sequences of the legume MLOS using the ProtTest 3.450. Subsequently, the ML tree was obtained on 1,000 bootstrap replicates using the JTT substitution model with ProtTest 3.450. Subsequently, the ML tree was obtained on 1,000 bootstrap replicates using the JTT substitution model with gamma distribution of 8 categories and α = 1.05 following the ProtTest estimation.

Protein characterization and motif prediction. The deduced amino acid sequences of the potential MLO genes were subjected to several prediction programs to determine their sub-cellular localizations28,51, protein topologies29–31,32 and to identify functional domains32. The prediction servers used in this study are listed in Supplemental Table S2. Except otherwise stated, the prediction server were run with default settings. The result of these predictions was then used to draw the protein organization of each potential MLO protein on the IBS server32 (Supplementary Table S2).

Conserved motifs were determined with the MEME algorithm30 (Supplementary Table S2). The MEME parameters were set to search for a maximum of 15 motifs with a motif width comprised between five and 50 residues. Presence or absence of the conserved motifs in each MLO sequence was then determined using FIMO and MAST algorithms also available from the MEME suite web server30 (Supplementary Table S2).

References
1. Rubiales, D. & Mikic, A. Introduction: legumes in sustainable agriculture. Crit. Rev. Plant. Sci. 34, 2–3, 10.1080/07352689.2014.897896 (2015).
2. Jensen, E. S. et al. Legumes for mitigation of climate change and the provision of feedstock for biofuels and biorefineries. A review. Agron. Sustain. Dev. 32, 329–364, 10.1007/s11353-011-0056-7 (2012).
3. Peix, A., Ramirez-Bahena, M. H., Velazquez, E. & Bedmar, E. J. Bacterial associations with legumes. Crit. Rev. Plant. Sci. 34, 17–42, 10.1080/07352689.2014.897899 (2015).
4. Rubiales, D. et al. Achievements and challenges in legume breeding for pest and disease resistance. Crit. Rev. Plant. Sci. 34, 195–236, 10.1080/07352689.2014.898445 (2015).
5. Hoekenga, R., et al. Cell biology of the plant–powdery mildew interaction. Curr. Opin. Plant Biol. 14, 738–746, 10.1016/j.pbi.2011.08.002 (2011).
| Page  | 41. Reinprecht, Y. | 45. Altschul, S. F. |
|-------|------------------|-------------------|
| 40. Cannon, S. B., May, G. D. & Jackson, S. A. Three sequenced legume genomes and many crop species: rich opportunities for... |
| 44. Lavin, M., Herendeen, P. S. & Wojciechowski, M. F. Evolutionary rates analysis of... | 10.1001/j.american.composition.2005.5098576 (2003) |
| 35. Panstruga, R. Discovery of novel conserved peptide domains by ortholog comparison within plant multi-protein families. | 10.1101/j.genetics.2006.531163 (2006) |
| 42. Liu, Q. & Zhu, H. Molecular evolution of the... | 10.1101/j.molecular.model.2006.531163 (2006) |
| 37. Feechan, A., Jermakow, A. M., Torregrosa, L., Panstruga, R. & Dry, I. B. Identification of grapevine... | 10.1101/j.molecular.model.2006.531163 (2006) |
| 31. Chen, Y., Wang, Y. & Zhang, H. Genome-wide analysis of the... | 10.1101/j.molecular.model.2006.531163 (2006) |
| 30. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME suite. | 10.1101/j.molecular.model.2006.531163 (2006) |
| 28. Nakai, K. & Horton, P. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. | 10.1101/j.molecular.model.2006.531163 (2006) |
| 25. Kessler, S. A. | 10.1101/j.molecular.model.2006.531163 (2006) |
| 22. Risk, J. M. | 10.1101/j.molecular.model.2006.531163 (2006) |
| 11. Bai, Y. | 10.1101/j.molecular.model.2006.531163 (2006) |
| 8. Pavan, S., Jacobsen, E., Visser, R. G. F. & Bai, Y. Loss of susceptibility as a novel breeding strategy for durable and broad-spectrum powdery mildew resistance... | 10.1101/j.molecular.model.2006.531163 (2006) |
| 7. Chauhan, H. | 10.1101/j.molecular.model.2006.531163 (2006) |
| 6. Lyngkjaer, M. F., Newton, A. C., Atzema, J. L. & Baker, S. J. The barley mlo-gene: an important powdery mildew resistance source. | 10.1101/j.molecular.model.2006.531163 (2006) |
| 5. Altschul, S. F. | 10.1101/j.molecular.model.2006.531163 (2006) |
| 4. Lavin, M., Herendeen, P. S. & Wojciechowski, M. F. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. | 10.1101/j.molecular.model.2006.531163 (2006) |
| 3. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. | 10.1101/j.molecular.model.2006.531163 (2006) |
| 2. Altschul, S. F. et al. Naturally occurring broad-spectrum powdery mildew resistance in a central American tomato accession is caused by loss of Mlo function. | 10.1101/j.molecular.model.2006.531163 (2006) |
| 1. Bai, Y. | 10.1101/j.molecular.model.2006.531163 (2006) |
| 46. | 10.1101/j.molecular.model.2006.531163 (2006) |
46. Solovyev, V., Kosarev, P., Seledsov, I. & Vorobyev, D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* 7 Suppl 1, S10, 10.1186/gb-2006-7-s1-s10 (2006).
47. Keese, M. *et al.* Gnomeous basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649, 10.1093/bioinformatics/bts199 (2012).
48. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729, 10.1093/molbev/msw178 (2013).
49. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552 (2000).
50. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27, 1164–1165, 10.1093/bioinformatics/btr088 (2011).
51. Chou, K.-C. & Shen, H.-B. Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. *Plos One* 5, e11335, 10.1371/journal.pone.0011335 (2010).
52. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786, 10.1038/nmeth.1701 (2011).
53. Liu, W. *et al.* IBS: an illustration for the presentation and visualization of biological sequences. *Bioinformatics* 31, 3359–3361, 10.1093/bioinformatics/btv362 (2015).
54. Young, N. D. *et al.* The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* 480, 520–524, 10.1038/nature10625 (2011).
55. Varshney, R. K. *et al.* Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* 31, 240–246, 10.1038/nbt.2491 (2013).
56. Yang, H. *et al.* Draft genome sequence, and a sequence-defined genetic linkage map of the legume crop species *Lupinus angustifolius* L. *Plos One* 8, e64799, 10.1371/journal.pone.0064799 (2013).
57. Bertoti, D. I. *et al.* The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat. Genet.* 48, 438–446, 10.1038/ng.3517 (2016).
58. Varshney, R. K. *et al.* Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat. Biotechnol.* 30, 83–89, 10.1038/nbt.2022 (2012).
59. Schmutz, J. *et al.* A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* 46, 707–713, 10.1038/ng.3008 (2014).
60. Kang, Y. J. *et al.* Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nat. Commun.* 5, 5443, 10.1038/ncomms5443 (2014).

Acknowledgements
This work was supported by the European project LEGATO (FP7-KBBE2013.1.2-02-613551), and the Spanish project AGL2014-52871-R from the Spanish Ministry of Economy and Competitiveness (MINECO) and co-financed by European fund for regional development (FEDER). NR is holder of a Ramón y Cajal postdoctoral position from the Spanish Ministry of Economy and Competitiveness.

Author Contributions
N.R. designed and performed the experiments, analysed and interpreted the data and wrote the manuscript; D.R. participated in the experimental designs, supervised the experiments and contributed in drafting and revising the manuscript. All authors read and approved the final manuscript.

Additional Information
Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Rispail, N. and Rubiales, D. Genome-wide identification and comparison of legume MLO gene family. *Sci. Rep.* 6, 32673; doi: 10.1038/srep32673 (2016).