Comprehensive Phylogenetic Analysis Sheds Light on the Diversity and Origin of the MLO Family of Integral Membrane Proteins

Stefan Kusch[1], Lina Pesch[1] and Ralph Panstruga[1,*]

[1]Unit of Plant Molecular Cell Biology, Institute for Biology I, RWTH Aachen University, 52056 Aachen, Germany

*Corresponding author: E-mail: panstruga@bio1.rwth-aachen.de.

Accepted: February 13, 2016

Abstract

Mildew resistance Locus O (MLO) proteins are polytopic integral membrane proteins that have long been considered as plant-specific and being primarily involved in plant–powdery mildew interactions. However, research in the past decade has revealed that MLO proteins diverged into a family with several clades whose members are associated with different physiological processes. We provide a largely increased dataset of MLO amino acid sequences, comprising nearly all major land plant lineages. Based on this comprehensive dataset, we defined seven phylogenetic clades and reconstructed the likely evolution of the MLO family in embryophytes. We further identified several MLO peptide motifs that are either conserved in all MLO proteins or confined to one or several clades, supporting the notion that clade-specific diversification of MLO functions is associated with particular sequence motifs. In baker's yeast, some of these motifs are functionally linked to transmembrane (TM) transport of organic molecules and ions. In addition, we attempted to define the evolutionary origin of the MLO family and found that MLO-like proteins with highly diverse membrane topologies are present in green algae, but also in the distinctly related red algae (Rhodophyta), Amoebozoa, and Chromalveolata. Finally, we discovered several instances of putative fusion events between MLO proteins and different kinds of proteins. Such Rosetta stone-type hybrid proteins might be instructive for future analysis of potential MLO functions. Our findings suggest that MLO is an ancient protein that possibly evolved in unicellular photosynthetic eukaryotes, and consolidated in land plants with a conserved topology, comprising seven TM domains and an intrinsically unstructured C-terminus.

Key words: MLO (Mildew resistance locus O), polytopic integral membrane protein, intrinsically disordered protein, functional divergence, Rosetta stone protein.

Introduction

Barley MLO (Mildew resistance Locus O) is the prototype of a family of sequence-diversified integral membrane proteins that, to date, have not been reported to occur outside the plant kingdom. The protein was first described in the context of resistance against powdery mildew infection. Loss-of-function mutations in the barley (*Hordeum vulgare*) *Mlo* gene confer durable broad-spectrum resistance against the fungal pathogen *Blumeria graminis* f.sp. *hordei*, which is the causal agent of the widespread plant disease (Jørgensen 1992; Büschges et al. 1997). Although originally discovered in barley, *mlo*-based resistance against powdery mildew is not confined to this cereal, but has been also found on the basis of natural or induced mutations in other monocotyledonous (monocot) and eudicotyledonous (eudicot) plant species (Consonni et al. 2006; Bai et al. 2008; Humphry et al. 2011; Wang et al. 2014). Accordingly, these plants also harbor *Mlo* genes, which are typically present as medium-sized gene families with ca. 10–20 members per species (Acevedo-Garcia et al. 2014). In the eudicot model species *Arabidopsis thaliana*, three of its altogether 15 *MLO* genes (*AtMLO2*, *AtMLO6*, and *AtMLO12*; note that the nomenclature in barley (*Mlo*) and Arabidopsis (*MLO*) differs) are the functional complements (co-orthologs) of barley *Mlo*. The three genes unevenly contribute to the powdery mildew resistance/susceptibility phenotype, with *AtMLO2* having a major role and *AtMLO6* and *AtMLO12* playing minor roles. Collectively, this phenomenon is also referred to as "unequal genetic redundancy" (Consonni et al. 2006).

© The Author 2016. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

MLO family members possess a seven-transmembrane (TM) domain topology with an extracellular (or luminal) amino terminus and an intracellular carboxyl terminus, resulting in three extracellular/luminal and three intracellular (cytoplasmic) loops (Devoto et al. 1999). Four invariant cysteine residues, which are located in two of the extracellular/luminal loops, are supposed to form two disulphide bridges that are crucial for MLO protein function and/or stability (Elliott et al. 2005). A number of additional highly conserved amino acids form a dispersed scaffold around which stretches with more variable amino acids can be found (Elliott et al. 2005). The second extracellular/luminal loop and the cytoplasmic C-terminus in particular are protein regions that exhibit high levels of sequence variability (Devoto et al. 1999, 2003). The proximal part of the intracellular carboxyl terminus nevertheless contains a binding site for the ubiquitous calcium sensor calmodulin that is conserved throughout the protein family (Kim et al. 2002a). At least in barley Mlo, the second and third cytoplasmic loops appear to be of special functional relevance, since known mutation-induced single amino acid substitutions that lead to a loss of Mlo function tend to cluster in these regions (Reinstädler et al. 2010). While barley Mlo accumulates in the plasma membrane (Devoto et al. 1999), the subcellular localization of Arabidopsis MLO4 appears to be less specific and involves a significant pool of endomembrane-localized protein (Chen et al. 2009).

The biochemical function of MLO proteins remains largely unexplored. Apart from the conserved calmodulin-binding domain (CaMBD), no further peptide motif has been identified that could provide a lead to its biochemical activity. Despite their characteristic membrane topology, which is reminiscent of metazoan G protein-coupled receptors, there is no experimental evidence that MLO proteins function *via* coupling to canonical (heterotrimeric) G proteins (Kim et al. 2002a; Chen et al. 2009; Lorek et al. 2013). Barley *Mlo* and Arabidopsis *MLO2* are expressed together with genes important for plant immunity (Humphry et al. 2010), which suggests a role for these MLO proteins in basal defense. Among the co-expressed genes, a considerable number codes for Receptor-like kinases (RLKs), a class of membrane receptors that play a pivotal role in plant immunity (Humphry et al. 2010). Genetic data and yeast-based protein–protein interaction studies also support a functional link of MLO proteins to RLKs (Kessler et al. 2010; Lalonde et al. 2010).

MLO proteins have not only been implicated in susceptibility/resistance against the powdery mildew disease. In addition to the resistance phenotype, barley *mlo* mutants show spontaneous accumulation of local callose deposits in the absence of any pathogen (Wolter et al. 1993) and deregulated cell death of leaf mesophyll cells (Peterhänsel et al. 1997). Similar phenotypes can be seen in Arabidopsis mutants with defects in either *MLO2* (*mlo2* single mutant) or *MLO2*, *MLO6*, and *MLO12* (*mlo2 mlo6 mlo12* triple mutant; Consonni et al. 2006). Furthermore, mutations in *MLO* genes can modulate processes other than susceptibility against the powdery mildew disease or inhibition of premature leaf senescence. *Atmlo4* and *Atmlo11* mutants show aberrant root thigmomorphogenesis, which is exemplified as unusual root curling upon a tactile stimulus under *in vitro* conditions. Although *AtMLO14* is phylogenetically closely related to *AtMLO4* and *AtMLO11*, the *Atmlo14* mutant does not display this growth behavior and also does not add to the phenotype in respective double or triple mutants (Chen et al. 2009; Bidzinski et al. 2014). Another example is *nortia* (*nta*, *Atmlo7*), a female gametophytic mutant that exhibits a pollen tube overgrowth phenotype in the synergids, that is, special cells of the plant's female fertilization apparatus, which ultimately results in reduced fertility (Kessler et al. 2010). In addition, rice *Os*MLO12 was shown to be required for pollen hydration, which occurs prior to pollen tube formation upon contact with an appropriate stigma (Yi et al. 2014).

Little is known about the phylogenetic history of MLO proteins. They occur in all higher land plant species analyzed to date and their origin dates at least back to the mosses some 400–450 million years ago (Ma; Devoto et al. 2003). In the course of evolution, the protein family diversified into several subfamilies, which can be recognized on the basis of phylogenetic analysis. Accordingly, at present MLO proteins have been categorized into seven phylogenetic clades (Jiwan et al. 2013; Zhou et al. 2013; Acevedo-Garcia et al. 2014). The above-mentioned MLO members, with defined loss-of-function phenotypes, group in clades IV and V (powdery mildew susceptibility/resistance), clade I (root thigmomorphogenesis), and clade III (pollen tube hydration and growth; Acevedo-Garcia et al. 2014). However, this categorization is being debated, as some studies imply the existence of four, six, or even eight MLO clades (Feechan et al. 2008; Deshmukh et al. 2014; Pessina et al. 2014). Moreover, the current phylogenetic analyses are incomplete since they do not include basal angiosperms, gymnosperms, or ferns.

Here, we revisited the MLO phylogeny to trace the potential evolutionary origin of this protein family. On the basis of comprehensive analyses, we were able to deduce a likely order in which the various embryophyte MLO clades evolved. This study further revealed conserved general and clade-specific MLO sequence motifs and disclosed the C-terminal tail of MLO proteins as being intrinsically disordered (ID). Careful inspection of genomes from non-land plant species led to the discovery of MLO-like proteins with highly diversified membrane topologies in several lineages, including green and red algae as well as chromalveolates. Rare fusions of MLO proteins with other types of proteins may provide an insight into potential MLO functions.

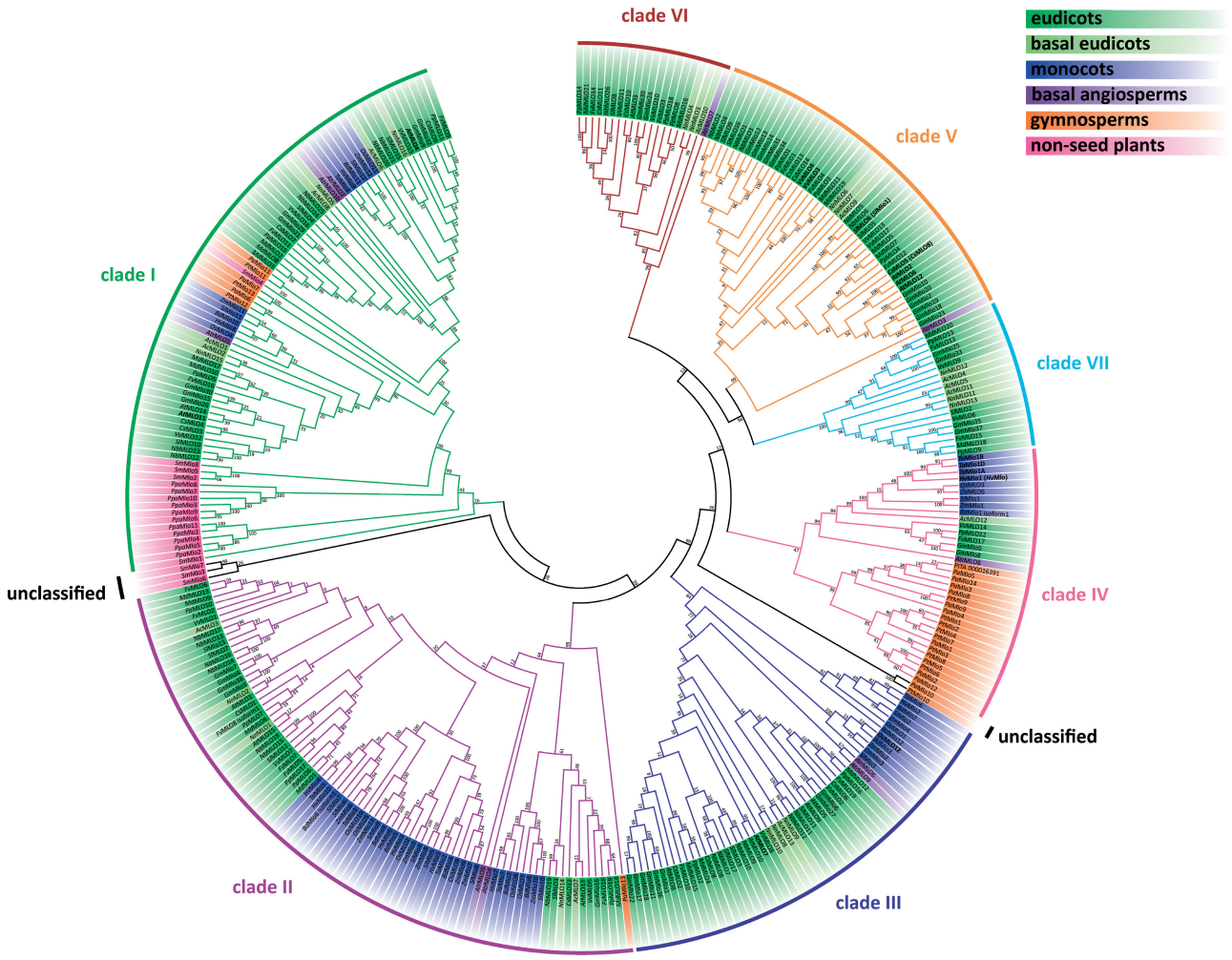

Fig. 1.—Maximum-Likelihood phylogenetic tree of land plant MLO proteins. Based on the total alignment of 341 MLO proteins (supplementary file 2, Supplementary Material online), the unrooted phylogenetic tree (cladogram) was calculated *via* the Maximum-Likelihood method, using JTT modeling with 1,000 bootstrap replications. Bootstrap values are indicated (in percent) at the base of each clade. The color code associated with the protein names indicates the affiliation of MLOs to particular taxonomic groups as indicated in the legend. MLO proteins labeled in bold font represent MLOs linked to physiological functions; that is, *At*MLO4 and *At*MLO11, root thigmomorphogenesis (Chen et al. 2009); *At*MLO7, pollen tube perception (Kessler et al. 2010); *Os*MLO12, pollen hydration (Yi et al. 2014); *Hv*Mlo, *Ta*Mlo1A, *Ta*Mlo1B, *Ta*Mlo1D, *At*MLO2, *At*MLO6, *At*MLO12, *Sl*MLO6 (syn. *Sl*Mlo1), *Cs*MLO5 (syn. *Csa*MLO8), powdery mildew interaction (Büschges et al. 1997; Consonni et al. 2006; Bai et al. 2008; Wang et al. 2014; Berg et al. 2015). The colored tree branches highlight clades I–VII. MLO proteins that did not sort in any of these clades were labeled "unclassified". Note that PITA_000016391 represents a *P. taeda* protein from which only a minor part aligns with MLO. This protein was therefore treated as pseudo-MLO and not included in further analysis.

Results

Identification and Annotation of MLO Proteins from Additional Land Plant Species

To date, published genome-wide MLO family information and phylogenetic MLO family analysis is restricted to some eudicot species, a few monocot species, the lycopod (spikemoss) *Selaginella moellendorffii* and the moss *Physcomitrella patens* (reviewed in Acevedo-Garcia et al. 2014). To considerably broaden this dataset and to cover further land plant (embryophyte) lineages, we performed extensive BLAST searches in the genomes of several additional species below. These include the eudicot *Nicotiana benthamiana* (*Nb*), the monocots *Brachypodium distachyon* (*Bd*), *H. vulgare* (*Hv*, barley), *Setaria italica* (*Si*), and *Zea mays* (*Zm*, maize), as well as two gymnosperm species, that is, *Picea abies* (*Pa*) and *Pinus taeda* (*Pt*). In addition, two basal eudicots, *Aquilegia coerulea* (*Ac*) and *Nelumbo nucifera* (*Nn*), and the basal angiosperm *Amborella trichopoda* (*Atr*) were included in the analysis. Finally, the genomes of *S. moellendorffii* (*Sm*) and *P. patens* (*Ppa*) were revisited to explore potentially updated genome information. Fern species could not be analyzed in this study, as no fern genome sequence has been released to date. For retrieval of novel MLO sequences, we used representative

Table 1
MLO Family Members of Selected Land Plant Species and Their Phylogenetic Classification

Plant species	Common name	Number of *MLO* genes	Phylogenetic clade								Reference
			I	II	III	IV	V	VI	VII	UC[a]	
Eudicots											
A. thaliana	Thale cress	15	3	3	5	–	3	1	–	–	Chen et al. (2006)
Cucumis sativus	Cucumber	14	4	2	3	–	3	2[b]	–	–	Zhou et al. (2013)
Fragaria vesca	Wild strawberry	19	3	6[b]	2[b]	1	3[b]	2[b]	2	–	Pessina et al. (2014)
Glycine max	Soybean	39	7	5	8	2	11	2[b]	4[b]	–	Deshmukh et al. (2014)
M. domestica	Apple tree	21	5	5	3	–	4	2	2	–	Pessina et al. (2014)
N. benthamiana	Tjuntiwari/Muntju	26	6	7	6	–	6	1	–	–	This study
Prunus persica	Peach	19	3	5	2	1	3	3	2	–	Pessina et al. (2014)
S. lycopersicum	Tomato	16[b]	3	4	3	–	4	1	1	–	Chen et al. (2014)
Vitis vinifera	Grapevine	17	3	3	2	1	4[b]	2	2[b]	–	Feechan et al. (2008)
Basal eudicots											
A. coerulea	Colorado blue columbine	13	4	2	1	1	1	1	3	–	This study
N. nucifera	Indian lotus	16	3	3	3	–	2	2	3	–	This study
Monocots											
Brachypodium distachyon	Purple false brome	12	2	6	3	1	–	–	–	–	This study
H. vulgare	Barley	11	2	7	1	1	–	–	–	–	This study
O. sativa	Asian rice	12	2	6	2[b]	2[b]	–	–	–	–	Liu and Zhu (2008)
S. italica	Foxtail millet	12	2	7	2	1	–	–	–	–	This study
Triticum aestivum	Bread wheat	7	1	4[b]	1[b]	1	–	–	–	–	Konishi et al. (2010)
Z. mays	Corn/Maize	13	2	7	3	1	–	–	–	–	This study
Basal angiosperms											
A. trichopoda	Amborella	10	3	2	2	1	1	1	–	–	This study
Gymnosperms											
P. abies	Norway Spruce	14	3	1	–	9	–	–	–	1	This study
P. taeda	Loblolly pine	13	3	–	–	9	–	–	–	1	This study
Lycopodiophyta											
S. moellendorffii	Spikemoss	8	5[b]	–	–	–	–	–	–	3[b]	This study
Mosses (Bryophyta)											
P. patens	–	11[b]	11[b]	–	–	–	–	–	–	–	This study

[a]Unclassified.
[b]Values differ from previously published numbers.

MLO proteins from all previously described phylogenetic clades of Arabidopsis, tomato (*Solanum lycopersicum*) and rice (*Oryza sativa*) as query sequences in BLAST searches. This procedure was chosen to maximize the likelihood that all MLOs of the respective species are discovered.

On the basis of our BLAST searches we identified in total 159 MLO proteins in the 12 species analyzed (table 1). Since the annotation of most of the genomes was incomplete we had to manually curate the gene models underlying the MLO proteins in many cases. The number of MLO proteins per newly analyzed species ranged from 8 (*S. moellendorfii*) to 26 (*N. benthamiana*), the latter being the second highest number of MLOs (after soybean with 39 members; Deshmukh et al. 2014) found in a plant species. Deviating from a former study (Jiwan et al. 2013), we detected eleven instead of eight MLOs for the moss *P. patens*. Taken together, we increased the number of validated MLOs identified in genome-wide searches substantially. These now comprise 338 unique proteins from 22 species, covering all major embryophyte lineages (except ferns). This dataset forms a solid basis for a comprehensive analysis of the phylogenetic relationships within and evolution of the embryophyte MLO family.

The Embryophyte MLO Family Splits into at Least Seven Phylogenetic Clades

We subjected all MLO protein sequences from previously published genome-wide searches and the MLOs newly identified in this study (table 1 and supplementary file 1, Supplementary Material online; 341 amino acid sequences in total, corresponding to 338 unique MLO protein sequences) to detailed phylogenetic analysis. The dataset included the three near-identical wheat homeologs *Ta*Mlo1A, 1B and 1D (Elliott et al. 2002), of which only *Ta*Mlo1A was included in further analysis. In addition, a sequence from *P. taeda* (designated PITA_000016391) represents a protein where only a minor part aligns with MLO. This protein was therefore treated as a pseudo-MLO and excluded from additional analysis.

First, the MLO amino acid sequences were aligned *via* ClustalW and the multiple sequence alignment was curated

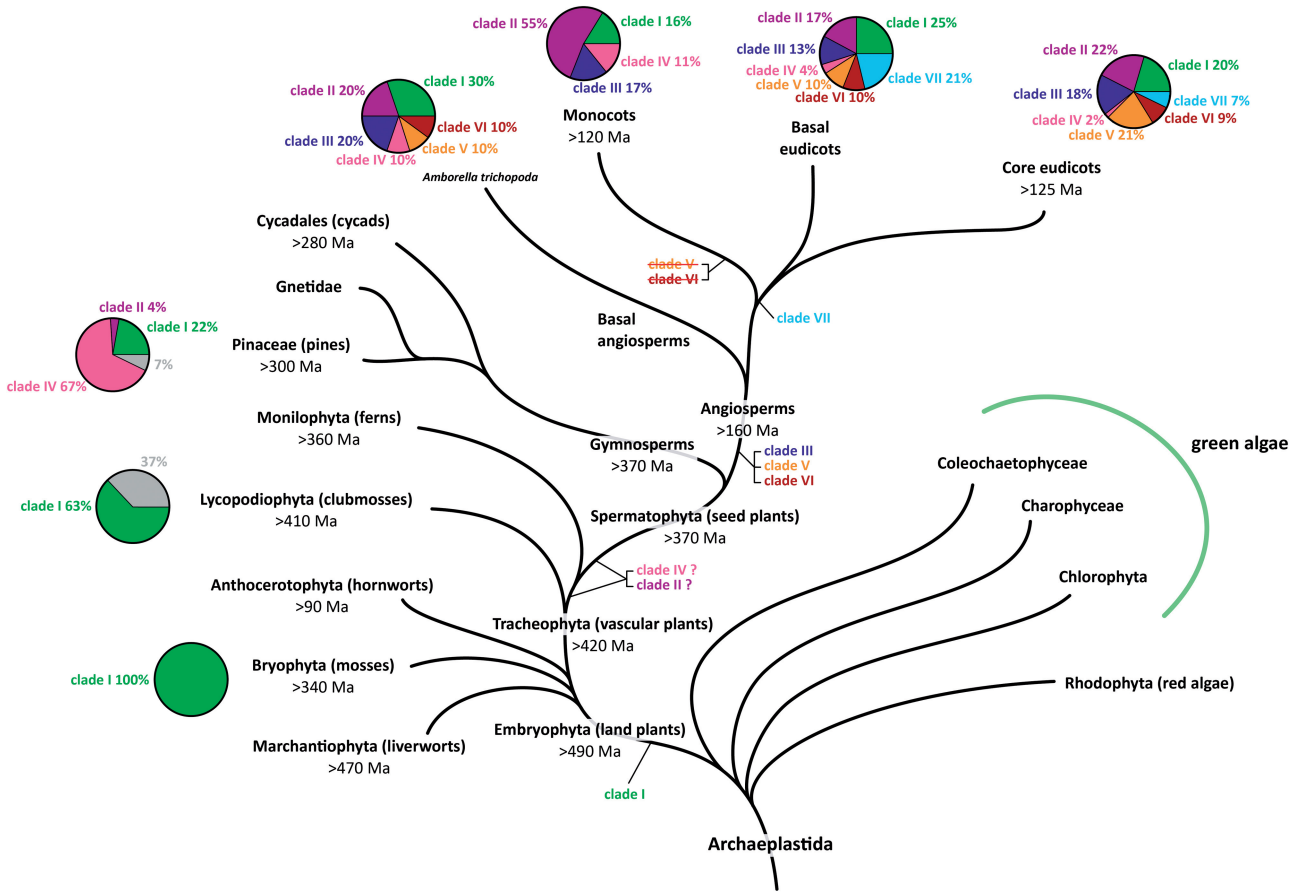

Fig. 2.—Inferred order of the appearance of MLO clades. The scheme depicts a simplified phylogeny of the major lineages within the Archaeplastida. The presumed order of appearance of the MLO clades, as inferred from the phylogenetic analysis in this study, is given on the respective plant clade branches. The minimum age of the main plant orders is given in million years ago (Ma). Pie charts illustrate the proportion of MLOs belonging to a given clade in the particular taxonomic group (unclassified MLOs shown in black).

manually (supplementary file 2, Supplementary Material online). This alignment was then used to generate phylogenetic trees with different algorithms (Maximum-Likelihood, Maximum parsimony, Neighbor-Joining, and UPGMA). These procedures yielded similar results; for clarity we focus here on the findings obtained with the Maximum-Likelihood method (fig. 2 and supplementary fig. 1, Supplementary Material online). This tree is comprised of seven clades, supporting the general relationship of MLO proteins revealed in most former studies (Jiwan et al. 2013; Zhou et al. 2013; Acevedo-Garcia et al. 2014). Three MLOs from *S. moellendorffii* and one MLO from *P. abies* and *P. taeda* each could not be unambiguously sorted into one of these clades (labeled unclassified).

The MLO proteins from the non-seed plants *S. moellendorfii* and *P. patens* all sort into clade I (fig. 1 and table 1). The MLOs identified in the pine species *P. abies* and *P. taeda* are found in clades I, II, and IV, but not in clade III. This could be interpreted as evidence that MLOs from clade III are not present in gymnosperms. However, it has to be considered that only one group of gymnosperm species (the *Pinaceae*) could be analyzed in this study. Interestingly, the basal angiosperm Amborella already harbors MLOs from clades I–VI. The basal eudicots *N. nucifera* and *A. coerulea* possess an additional clade VII, as do six of the nine eudicot species. In contrast to Amborella and the eudicots, all monocots only harbor MLOs from clades I–IV, but no MLOs that belong to clades V, VI, or VII. In turn, 6 of the 11 eudicot species from this study do not possess an MLO from clade IV.

To analyze this complex situation in detail, we generated a scheme representing the evolutionary history of the kingdom Plantae (fig. 2), labeling the deduced approximate appearance (or disappearance) of the respective MLO clades in the course of plant evolution. As mentioned before, clade I appears to be the most ancient MLO clade, dating back at least 490 Ma to the oldest known land plants (Becker and Marin 2009). Next, clades II and IV seem to have evolved, as these are present in gymnosperms, whereas clade III is seemingly absent in this

lineage. Early in angiosperm evolution clades III, V, and VI seem to have evolved, since Amborella has MLOs from clades I–VI. As a basal angiosperm Amborella is thought to have divided from the other angiosperms very early (>160 Ma), before the split of monocots and eudicots (ca. 120 Ma). Therefore, the ancestor of the monocots may have harbored MLOs from clades I–VI as well, but during monocot evolution, members of clades V and VI were seemingly lost. Eudicots, however, generally possess MLOs from clades I–VII, though some of them have lost MLOs from clade IV or VII during diversification of the eudicot lineage.

We noted an apparent over-representation of clade IV MLOs in the two gymnosperms: approximately three quarters of the MLOs found in these species group into this clade. Clade II, on the other hand, is under-represented, indicated by the fact that only one pine MLO (*Pa*Mlo13) was found to belong to this clade. Conversely, clade II MLOs are over-represented in monocot species, as > 50% of the monocot MLOs were classified into this clade. In eudicots, clade IV is under-represented and even lost in some species, suggesting that this clade plays a minor role in eudicots. It is possible that clades V–VII partially replaced clade IV in the course of evolution.

Expansion of the MLO Family in Eudicots

We noted that apart from eudicots the number of *MLO* genes per species ranges from 7 (wheat) to 14 (*P. abies*) in all land plant species, with a mean of 11.2 and a median of 12 MLOs/species (supplementary table 1, Supplementary Material online). Eudicot plant species, however, harbor between 13 (*A. coerulea*) and 39 (soybean) different *MLO* genes, with a mean of 19.5 and a median of 17 MLOs/species. This difference is statistically highly significant, with $P < 0.001$ when comparing eudicots with all other land plants (supplementary fig. 2, Supplementary Material online), prompting further examination. We compared the number of *MLO* genes with genome size, gene number, gene density (genes/Mbp genome size), and chromosome number (table 2) to determine whether the increase in the quantity of *MLO* genes may be due to genome expansion or duplication events, which have occurred several times in land plant evolution (Jiao et al. 2011). We found a strong positive correlation (Spearman's rank correlation method) between the number of *MLO* genes and gene number and chromosome number, respectively, for eudicots, but not for monocots (supplementary fig. 3*A* and *B* and supplementary table 2, Supplementary Material online). These correlations were statistically significant ($P < 0.01$). With regard to genome size and gene density, no correlation and no statistically significant difference were found (supplementary fig. 3*C* and *D*, Supplementary Material online). In addition, the largest genomes in this analysis were from wheat (a monocot species) and the two pines *P. abies* and *P. taeda*; hence the increase in *MLO* gene number cannot be explained by genome size expansion. An important feature of eudicot MLO families is the presence of three additional clades (V–VII) that are not present in other land plants except Amborella. Interestingly, when excluding clades V–VII MLOs from the calculation for eudicots, the mean drops to 12.1 and the median to 11 *MLO* genes/species in eudicots. These numbers are similar to those of the other land plants (see above and supplementary fig. 2, Supplementary Material online), suggesting that the elevated number of *MLO* genes in eudicot species is primarily based on the evolution of MLO clades V–VII rather than the consequence of genome duplication events during plant evolution.

MLO Proteins Harbor Conserved Motifs Associated with TM Transport

We conducted a more detailed analysis of protein motifs in the MLO proteins, based on the above-mentioned sequence alignment (supplementary file 2, Supplementary Material online). To this end, we used WebLogo-based motif representation of the more conserved regions (highlighted in supplementary fig. 4, Supplementary Material online) and performed MEME-based motif search on the poorly conserved first extracellular/luminal loop and the intracellular C-terminal region of the MLO protein. TM domains were excluded from this analysis, as these are mainly composed of aliphatic amino acids to facilitate membrane localization and therefore not informative.

First, we reduced the aligned sequences to a minimum consensus by eliminating all gapped regions (supplementary fig. 4), that is, parts of the alignment where > 95% of the MLO sequences had gaps owing to insertions in individual MLOs. Given that the overall consensus was quite low for certain regions, but high for these sections within single clades (especially in the first extracellular/luminal loop and the C-terminal cytoplasmic domain), the minimum consensus was generated for each clade separately. The clade-specific consensus sequences were then compared in another alignment (supplementary fig. 5 and supplementary file 3, Supplementary Material online), which is 557 amino acids in length. The positions in this alignment are henceforth used as the basis for positioning of conserved amino acids and motifs.

Next, we analyzed the aligned protein sequences for highly conserved amino acids, that is, amino acids that are present in > 90% of the MLO proteins at a particular position. In line with previous characterizations of the MLO protein family (Devoto et al. 2003; Elliott et al. 2005), we found four highly conserved cysteines (C^{83}, C^{96}, C^{129}, C^{386}) in the extracellular/luminal loops 1 and 3 (fig. 3). These may be involved in disulfide bridge formation and therefore be important for structural integrity of the protein (Elliott et al. 2005). We further corroborated the high conservation of a tryptophan residue (W^{442}) in the C-terminal intracellular region, which is known to be a key amino acid of the CaMBD (fig. 3; Kim

Table 2

Genomic Features of the Land Plant Species Considered in This Study

Plant species	Number of MLO genes	Genome size (Mbp)	Number of genes	Gene density (genes/Mbp)	Chromosome number (n)	Reference
Eudicots						
A. thaliana	15	135	27,416	203.1	5	Arabidopsis Genome Initiative (2000) and Lamesch et al. (2012)
Cucumis sativus	14	367	26,682	72.7	7	Huang et al. (2009)
Fragaria vesca	19	240	34,809	145.0	7	Shulaev et al. (2011)
Glycine max	39	1,115	46,430	41.6	20	Schmutz et al. (2010)
M. domestica	21	742.3	57,386	77.3	17	Velasco et al. (2010)
N. benthamiana	26	3,000	49,818	16.6	19	Bombarely et al. (2012)
Prunus persica	19	265	27,852	105.1	8	Verde et al. (2013)
S. lycopersicum	16	900	34,727	38.6	12	Tomato Genome Consortium (2012)
Vitis vinifera	17	487	26,346	54.1	19	Jaillon et al. (2007)
Basal eudicots						
A. coerulea	13	302	24,823	82.2	7	Unpublished*
N. nucifera	16	929	26,685	28.7	8	Ming et al. (2013)
Monocots						
Brachypodium distachyon	12	272	26,552	97.6	5	International Brachypodium Initiative (2010)
H. vulgare	11	5,100	30,400	6.0	7	Mayer et al. (2012)
O. sativa	12	372	39,049	105.0	12	Goff et al. (2002)
S. italica	12	515	35,471	68.9	9	Bennetzen et al. (2012)
Triticum aestivum	7	17,000	124,201	7.3	21	International Wheat Genome Sequencing Consortium (2014)
Z. mays	13	2,300	39,656	17.2	10	Schnable et al. (2009)
Basal angiosperms						
A. trichopoda	10	870	26,846	30.9	13	Amborella Genome Project (2013)
Gymnosperms						
P. abies	14	20,000	28,354	1.4	12	Nystedt et al. (2013)
P. taeda	13	20,150	50,172	2.5	12	Neale et al. (2014)
Lycopodiophyta						
S. moellendorffii	8	106	22,273	210.1	27	Banks et al. (2011)
Mosses (Bryophyta)						
P. patens	11	480	32,272	67.2	27	Rensing et al. (2008)

*These sequence data were produced by the US Department of Energy Joint Genome Institute, *A. coerulea* Genome Sequencing Project, http://www.phytozome.net/. (last accessed March 1, 2016)

et al. 2002a, 2002b). In addition, several conserved amino acids and short motifs were found, preferentially in the second and third cytoplasmic loops. These include a WXXWE motif (amino acids (aa) 175–179), ETXTLE (aa 181–186) a [CS]FX[RK]QF (aa 224–229) motif, DY (aa 237–238), a [LM]RXGF motif (aa 241–245), a FDF motif (aa 257–259), a FWFXXP motif (aa 348–353), two additional phenylalanines (F^{259}, F^{271}), an arginine (R^{265}), and a tyrosine (Y^{262}; fig. 3). Using the Find Individual Motif Occurrences (FIMO) tool from the MEME suite (Grant et al. 2011), we identified >1000 baker's yeast (*Saccharomyces cerevisiae*) proteins carrying one of these motifs or derivatives thereof (supplementary file 4, Supplementary Material online). The *S. cerevisiae* proteome database was used since the baker's yeast has one of the best-annotated genomes to date, hence facilitating detailed and comprehensive Gene Ontology (GO) analysis. The GO analysis revealed that the process "TM transport" was over-represented among the GO terms with the FWF sub-motif (table 3 and supplementary file 4, Supplementary Material online), suggesting a function in translocation across membranes. Additional terms include "carbohydrate metabolism" (FDF sub-motif) and "response to stress" (ETXTLE sub motif). No GO terms could be associated with the other motifs by this type of analysis.

We then compared the conserved regions between the single clades. Therefore, we used WebLogo-based motif representation for each clade (supplementary fig. 6). By this analysis, we noted signs of clade-specific diversification of certain MLO motifs. For example, several motifs were found to be conserved in most MLO clades except clades I and II, including $MA[E]_n$- (mostly clade I MLOs) or $MA[G]_n$-like motifs at the start of the MLO sequences, a lysine-arginine-rich motif (aa 44–49) in the first cytoplasmic loop that is interspersed with threonine (T), asparagine (N), or glutamine (Q) in clades I and

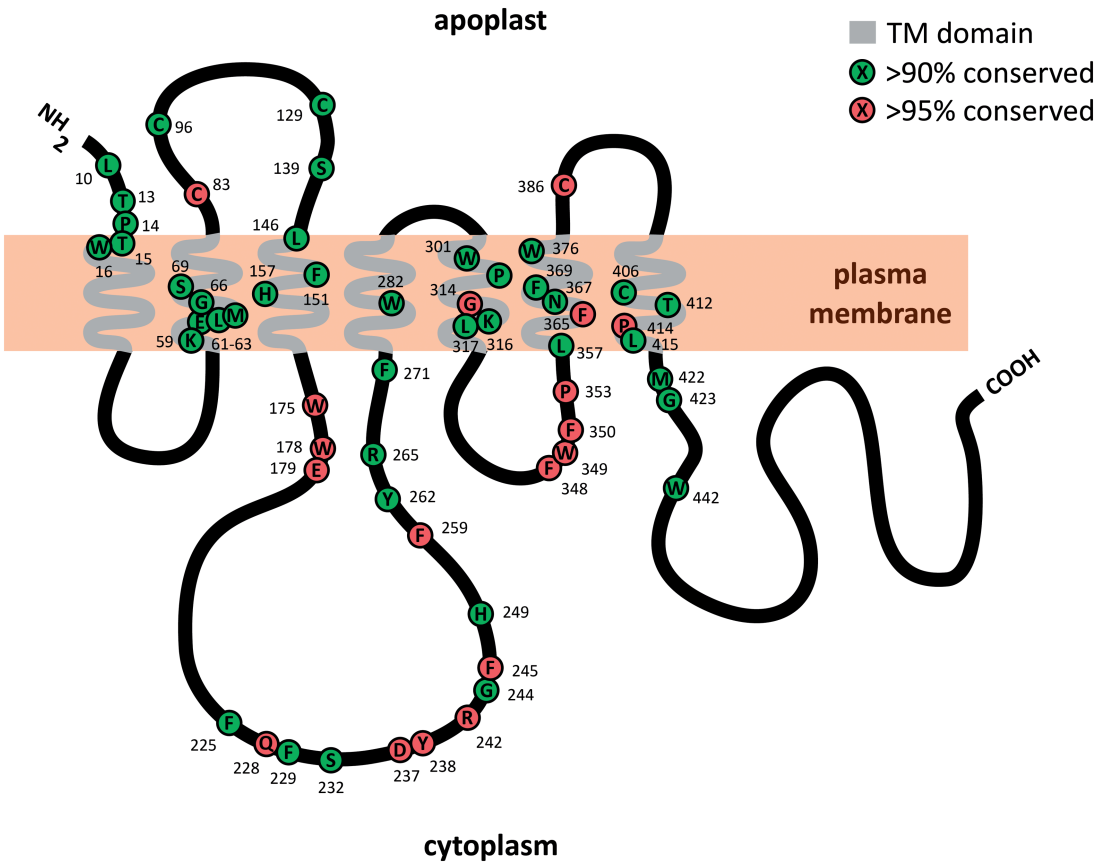

FIG. 3.—Conserved amino acids in embryophyte MLO proteins. The scheme depicts the MLO seven TM domain topology (based on Devoto et al. 1999). Amino acids that are conserved in ≥90% of all land plant MLOs are indicated in green, amino acids with ≥95% conservation (quasi-invariant) are shown in red (supplementary fig. 5, Supplementary Material online).

Table 3

Conserved Motifs in Land Plant MLO Proteins and Associated GO Terms

Motif	Location (aa)	Domain	Clades	Top GO terms
TPTW	13–16	TM1	All	Transposition, viral life cycle, RNA/DNA process
[FY]EALEK	52–57	Cytoplasmic loop 1	All	Fatty acid metabolism
RRLL	117–120	Apoplastic/luminal loop 1	I–V	None
WXXWE	175–179	Cytoplasmic loop 2	All	None
ETXTLE	181–186	Cytoplasmic loop 2	IV–VII	Cellular process, response to stress, regulation
RFR	191–193	Cytoplasmic loop 2	I, III–VII	Organic acid transport, ion transport
TSF[GV][RK]RH	199–205	Cytoplasmic loop 2	I, III–VII	Cellular process, hexose/carbohydrate transport
CFFRQF/SFFKQF	224–229	Cytoplasmic loop 2	All	None
TLR[LH]GFI	240–246	Cytoplasmic loop 2	All	None
[FY]DF	257–259	Cytoplasmic loop 2	All	Cellular process, carbohydrate metabolism
FWF	348–350	Cytoplasmic loop 3	All	Ion transport, TM transport
FG[IL][RK]SCF	381–387	Apoplastic/luminal loop 3	All	Transport
QHEI	542–545	Cytoplasmic tail	I–VI	Metabolic process, regulation, response to stimulus
[DE]FSF	550–553	Cytoplasmic tail	III, IV, V	None

II, and the motif XRFRXXTHXTSF[GV][RK]RH in cytoplasmic loop 2, present like this in clades III, IV, and VII, slightly different also in clades V and VI, and strongly diversified in clades I and II. GO analysis against the *S. cerevisiae* proteome revealed that both the RFR and the TSF[GV][RK]RH sub-motifs are associated with transport of amino acids, organic compounds, sugars, or anions (table 3, supplementary file 4, Supplementary Material online).

Next, we analyzed the non-conserved regions of the single clades, in particular the first extracellular/luminal loop (aa 98–126) and the C-terminal cytoplasmic domain (aa 451–557) using the MEME suite. In the second extracellular/luminal loop, we recognized a tetrapeptide motif (RRLL, aa 117–120) that is present in MLOs from all clades except clades VI and VII. The already described tetrapeptide motif [DE]FSF (Panstruga 2005) in the C-terminal tail appears in clades III, IV, and V, but cannot be linked to any GO terms (supplementary files 5 and 6, Supplementary Material online). Finally, the motif QHEI was identified in clade V, which is associated with "response to stimulus or stress" in yeast.

The C-Terminus of MLO Proteins Is Predicted to Be Intrinsically Unstructured

The proximal C-terminal region of MLO proteins is highly sequence-diversified (Devoto et al. 1999, 2003). In light of this, we hypothesized that the MLO C-terminus might be intrinsically unstructured, that is, lacking stable secondary structures such as α-helices or β-sheets. Many ID regions are known to undergo induced folding upon interaction with their physiological interaction targets, which associates ID regions with the unusual combination of low affinity and high specificity for binding partners. Hence, such interactions can be readily reversible. In addition, ID regions can confer flexibility and promiscuity to target binding (Dyson and Wright 2005). Furthermore, ID protein sequences often evolve at a faster rate than ordered protein sequences, likely due to a lack of structural constraints, resulting in a greater number of single amino acid substitutions, insertions, deletions, and repeat unit expansions (Brown et al. 2011; Nilsson et al. 2011). Since the excessive sequence diversity of the MLO C-terminus could be explained by such a scenario, we subjected all MLO protein sequences to disorder tendency analysis *via* IUPred (Dosztányi et al. 2005), a tool that predicts protein disorder based on estimated energy content of the protein region. Most MLO proteins, representing all clades, showed signs of disorder at their C-termini, predominantly at the distal ends, with only a few exceptions such as *At*MLO9 and *Ta*Mlo4, though *At*MLO9 has an extremely short C-terminus (supplementary file 7, Supplementary Material online). The tendency for an ID C-terminus appears to be an ancient and evolutionarily conserved feature, since most of the *Selaginella*, *Physcomitrella*, and gymnosperm MLOs have a predicted ID region in the cytoplasmic tail as well. A recent study conducting a codon-based evolutionary analysis of clade V *MLO* genes revealed that codons for both the first extracellular loop and the C-terminal tail are under strong positive (diversifying) selection, supporting the notion that these domains may be involved in protein–protein interactions (Iovieno et al. 2015). Together, these findings suggest that the C-terminus of MLO proteins comprises a functionally conserved, though sequentially diversified, ID domain that is possibly involved in specific binding to target proteins.

MLO Can Be Traced Back to Unicellular Eukaryotes

Since MLO is an evolutionarily ancient protein present in all land plants, we wondered how far MLO can be traced back in evolution. To this end, we performed extensive BLASTP searches against the available proteomes from different algae, bacteria, fungi, unicellular eukaryotes, and animals. We discovered many MLO-like proteins in Chlorophyta (green algae) species (*e*-values in the range of 4e-06 to 4e-45; supplementary file 8, Supplementary Material online), though not all green algae seem to carry *MLO*-type genes. For example, no MLO-like proteins were found in the fully sequenced species *Coccomyxa subellipsoidea* and *Dunaliella salina*, or in any algae from the family *Ulvophyceae*. However, representatives of the main families *Chlorophyceae*, *Trebouxiophyceae*, *Chlorodendrophyceae*, and *Prasinophyceae* (all Chlorophyta) harbor several MLO-like proteins, suggesting the presence of an MLO progenitor in the common ancestor of green algae and land plants. Amongst these, *Volvox carteri* (*Chlorophyceae*), a colony-forming alga, is the only multicellular representative, while all other species are unicellular green algae. In addition to green algae, MLO-like proteins can be found in a number of diverse phyla (BLASTP *e*-values >1e-06; fig. 4 and supplementary file 8, Supplementary Material online). We identified proteins with MLO-like amino acid sequences in diverse Chromalveolata species, including *Perkinsus* and *Eimeria* species (Alveolata, mostly animal pathogens), the brown alga *Ectocarpus siliculosus* (Heterokontophyta), and the ciliate *Tetrahymena thermophila*. Further, several oomycete species, that is, *Phytophtora infestans*, *Hyaloperonospora arabidopsidis*, *Albugo candida*, *A. laibachii*, *Phytium ultimum* (all of them plant pathogens), as well as *Aphanomyces invadans* and *Saprolegnia* sp. (animal pathogens) harbor one or several MLO-like proteins. Also, the red alga *Chondrus crispus* (Rhodophyta), the Rhizaria species *Plasmodiophora brassicae* (a plant pathogen), the Apusozoa *Thecamonas trahens* (flagellate protozoa) and the Amoebozoa species *Entamoeba histolytica* appear to possess proteins with similarity to MLO. No convincing MLO-like proteins (*e*-values >1) were found in bacteria, archaea, fungi, and multicellular animals (metazoa).

As with land plant MLOs, many of the MLOs from non-land plant organisms comprise six to eight predicted TM domains, though several proteins deviate markedly from this topology. In total, we discovered 14 different protein architectures with 2–21 predicted TM domains (supplementary files 8 and 9, Supplementary Material online). Reflecting the situation of embryophyte MLOs, most of these topologies include a predicted extracellular/luminal N-terminus and a cytoplasmic C-terminus. The extent of sequence similarity between land

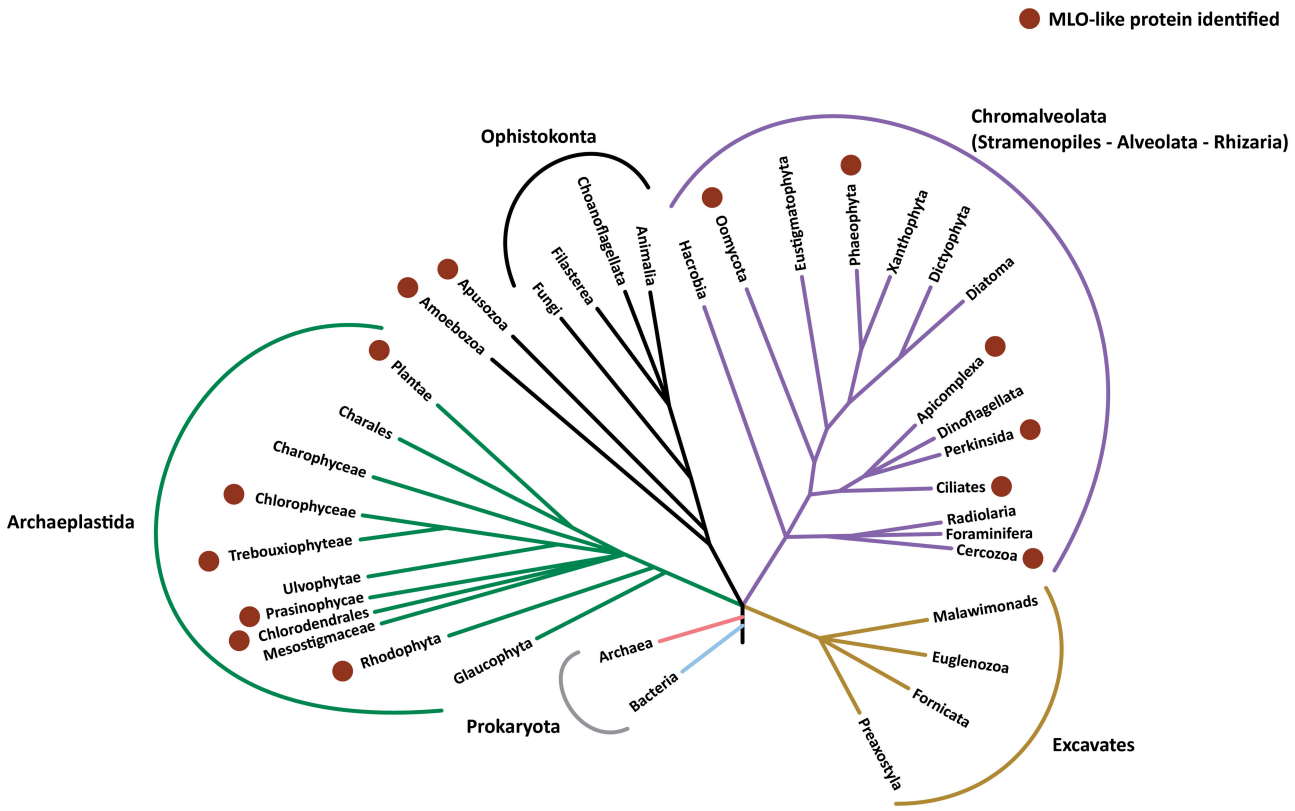

Fig. 4.—Occurrence of species with MLO and MLO-like proteins within the tree of life. The scheme depicts a simplified tree of life, indicating the kingdoms of life and many major clades within these kingdoms. The red dots indicate that at least one MLO-like protein was identified in a species belonging to the respective clade.

plant MLOs and the MLOs from non-land plant species also varies considerably. It ranges from a near-complete alignment (TM1–TM7) of the amino acid sequences to only short regions matching land plant MLO sequences, which in most cases at least comprises the stretch between TM1–TM3 (supplementary file 8, Supplementary Material online and fig. 5). Therefore, the region between TM1 and TM3 seems to define the least common denominator of MLO sequences from the various lineages and may indicate a core segment from which all MLO proteins evolved. The C-terminal part downstream of the last predicted TM domain is in all instances highly diversified and does not align to land plant MLO sequences.

MLO(-like) Proteins May Occur as Fusion Proteins in Rare Cases

In the context of the analysis of non-land plant MLOs, we discovered a number of MLO-like proteins that are predicted to be C-terminally fused with calmodulin-like (EF-hand rich) domains, which are generally implicated in Ca^{2+} ion sensing (Blanchard et al. 1997; Gifford et al. 2007). We noted that this type of hybrid protein occurred in several distantly related eukaryotes, such as green algae, oomycetes, and Apusozoa. Driven by this initial finding, we performed a comprehensive domain organization/architecture search with NCBI CDART and InterPro using Arabidopsis MLO4 (an ancient clade I MLO protein with defined function) as the query. We identified several additional predicted fusions of MLO proteins with other proteins/domains in the database, which are all exclusively present in land plant species (supplementary files 8 and 10, Supplementary Material online and fig. 6). These include a putative fusion with a glycosyl hydrolase domain in rosids (e.g., *V. vinifera* and *Populus euphratica*) and a potential fusion with an RNase-H-like domain in two Rosales, *Malus domestica* and *Morus notabilis* (both as fusions to the N-terminus of the MLO protein). In addition, fusions with a late embryogenesis abundant (LEA) protein domain in *N. nucifera* and with a mechanosensitive channel in proteobacteria were predicted.

Conservation of Sequence Patterns Known from Angiosperm MLOs

Next, we aimed at finding out whether highly conserved sequence patterns of angiosperm MLOs are present in all land plants and also in more distantly related eukaryotic MLO

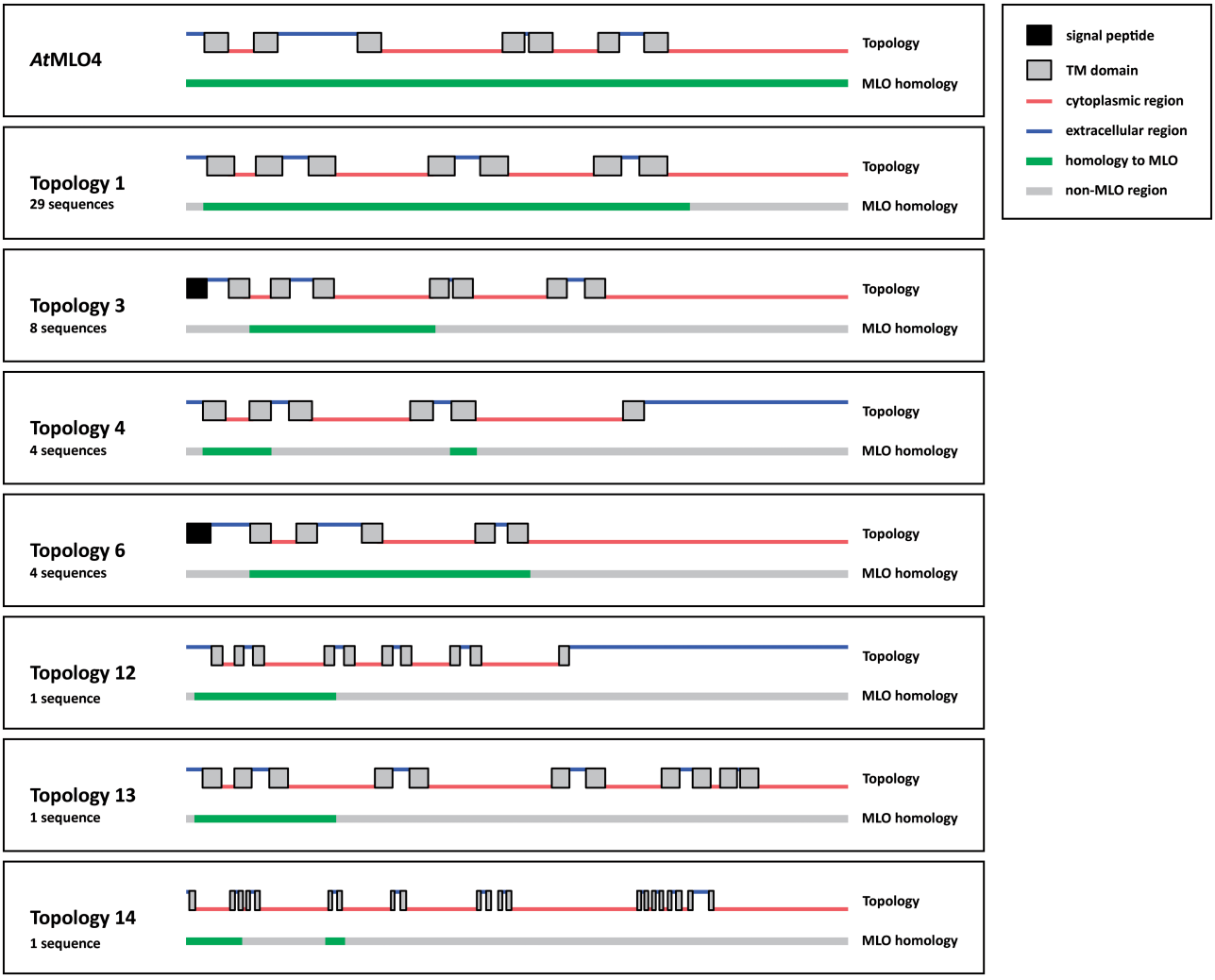

Fig. 5.—Predicted topologies of MLO-like proteins from non-land plant species. The schemes depict the topologies of MLO-like proteins identified in non-land plant species, as inferred by TOPCONS analysis (Tsirigos et al. 2015). The four most common topologies and three cases with relatively high sequence conservation to land plant MLOs (out of 14 different topologies in total) are shown. The region of recognizable sequence similarity to angiosperm MLO sequences is indicated by the green bar below the respective topology. Additional protein features are designated in the legend. The schemes represent the output graphs of TOPCONS. Note that the topologies are not shown to scale.

proteins. First, we analyzed gymnosperm, *Selaginella* and *Physcomitrella* MLOs for the presence of the four conserved cysteines (C^{83}, C^{96}, C^{129}, and C^{386}; Devoto et al. 1999). We found that the three cysteines in the first extracellular/luminal loop (C^{83}, C^{96}, and C^{129}) as well as the cysteine in the third apoplastic/luminal loop (C^{386}) are present in all of these MLOs (supplementary file 2, Supplementary Material online; see positions 270, 283, 422, and 830 in the alignment). In addition, the conserved CaMBD, which is situated at the proximal end of the cytoplasmic tail and includes the quasi-invariant W^{442}, is present in 16 of the overall 19 *Physcomitrella* and *Selaginella* MLO proteins. The exceptions are *Sm*Mlo3 and *Sm*Mlo6 (for which the sequence ends just after the seventh TM domain), and *Sm*Mlo7, which seems to lack W^{442}.

Then, we searched the MLO-like proteins from green algae, Alveolata, Stramenopiles, and other eukaryotic classes for the presence of these sequence features. These MLO-like proteins are difficult to align due to their high degree of sequence diversity. We succeeded in aligning a sub-region of the MLO-like proteins from green algae, comprising the region from TM1 to TM3 of the consensus MLO sequence (supplementary file 11, Supplementary Material online; until position 280 of the alignment). The cysteines C^{83} and C^{139} were both found in 94% of the MLO-like proteins from green algae. For C^{96}, the situation is less clear, as only 44% (8 of 18) of the sequences possessed a cysteine at this position. For the identification of C^{386} equivalents, we searched for SCF-like motifs in the predicted extracellular/luminal domains of the MLO-like

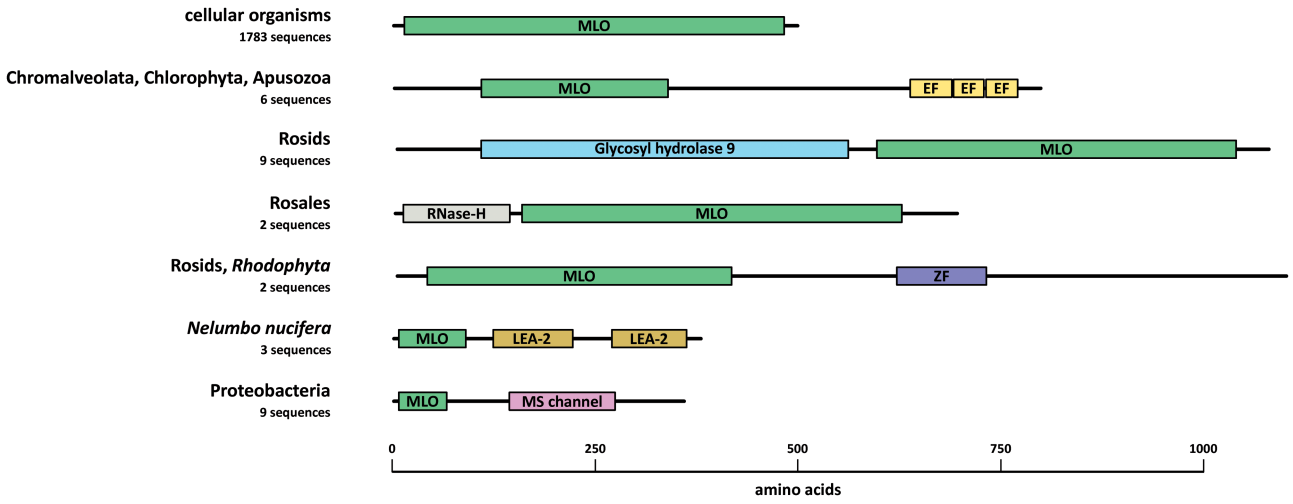

Fig. 6.—Potential MLO Rosetta stone proteins. Domain architecture searches with *At*MLO4 as a query in NCBI CDART (http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi [last accessed March 1, 2016], Geer et al. 2002) yielded instances where MLO proteins are seemingly fused with other proteins (supplementary file 10, Supplementary Material online). The schemes depict representative MLO fusions based on the output of the CDART search. The number of sequences indicates the quantity of independent occurrences of this domain architecture identified by the CDART tool.

proteins, which were present in seven of the analyzed sequences (39%; supplementary file 12, Supplementary Material online).

For the more distantly related MLO-like proteins from other eukaryotes the situation is even more complex. The MLO-like protein of the red alga *C. crispus*, the next relative to green algae, does not seem to possess any of the conserved cysteines. Given that no useful alignment was possible for most of the other eukaryotic MLO-like proteins, we searched the predicted extracellular/luminal domains for cysteines embedded in similar patterns as the conserved cysteines (ICI for C^{83}, PCX for C^{96}, and SCF for C^{386}; no such pattern was obvious for C^{139}). Few such cysteines were found (12 in total), seemingly in random order, in contrast to Archaeplastida MLOs (supplementary file 12, Supplementary Material online). Nine of these cysteines are embedded in patterns similar to that of C^{83}; one cysteine corresponds to the C^{96} pattern, and two cysteines correspond to the C^{386} pattern. The search for the conserved calmodulin domain revealed only a few vaguely similar sequences in both green algae and the more distantly related eukaryotes (supplementary file 13, Supplementary Material online), which are very diverse from the land plant MLOs and hence likely not directly related to the MLO CaMBD.

Taken together, our analysis shows that some of the cysteines (C^{83} and possibly C^{139}) are rather conserved in MLO-like proteins from non-angiosperms, green algae and to some extent even from some distantly related eukaryotes, whereas the other cysteines seem less conserved. The higher degree of conservation of C^{83} and C^{139} may indicate that these two residues form an ancestral pair of cysteines that possibly enabled a first apoplastic/luminal disulphide bridge to be formed. The second pair of cysteines (represented by C^{96} and C^{386}) seemingly evolved later, perhaps within the first land plants. Functional analysis on the basis of complementation efficiencies involving site-directed mutagenesis of these cysteines further support this grouping and suggest a more decisive functional role for C^{83} and C^{139} (Elliott et al. 2005). In contrast to the cysteines, the CaMBD is only conserved in non-angiosperm land plants but seems to be absent or highly diversified in algae and more distantly related eukaryotes.

Discussion

The MLO Protein Clades Likely Represent Functionally Distinct Entities

We provide a large dataset of MLO protein sequences from various land plant clades, including nine eudicots, two basal eudicots, six monocots, one basal angiosperm, two gymnosperms, a lycopod (clubmoss) and a moss (table 1 and supplementary file 1, Supplementary Material online). The dataset lacks MLO sequences from true ferns, as fern genomes have not been released yet (Sessa et al. 2014). Our phylogenetic analysis on the basis of this comprehensive dataset and using a variety of phylogenetic methods (Maximum-Likelihood, Maximum parsimony, Neighbor-Joining, and UPGMA) supports the view that land plant MLOs divide into seven clades (fig. 1 and supplementary fig. 1, Supplementary Material online; Jiwan et al. 2013; Zhou et al. 2013; Acevedo-Garcia et al. 2014), and not into six clades (Chen et al. 2006; Feechan et al. 2008) or eight clades (Pessina et al. 2014), as reported by others. However, three MLOs from *S. moellendorfii*, one *P. abies* MLO, and one *P. taeda* MLO were not sorted in any of the seven clades. This could indicate that gymnosperms possess an additional MLO clade in close relationship to

clades III and IV, and that clubmosses might harbor an extra clade closely related to clade I. Alternatively, the employed algorithms may not be powerful enough to group these MLOs into the appropriate clade with the present dataset, which includes relatively few non-seed plant and gymnosperm MLOs. In former studies (Jiwan et al. 2013), the *Selaginella* MLOs were all classified as clade I MLOs by consensus Bayesian phylogeny, rather supporting the latter argument.

Our analyses strengthen the notion that clade I is indeed the most ancient embryophyte MLO clade (Jiwan et al. 2013), as (a) all species in this analysis harbor one or several clade I MLOs and (b) all *Selaginella* and *Physcomitrella* MLOs group into this clade. Thus far, members of clade I have been associated with the regulation of root growth patterns, as *At*MLO4 and *At*MLO11 were shown to be required for proper root thigmomorphogenesis (Chen et al. 2009; Bidzinski et al. 2014). It is tempting to speculate that the early evolution of embryophyte MLOs is linked to the emergence of roots. However, two observations argue against this hypothesis: first, no root phenotype was found for the other clade I MLO in Arabidopsis (*At*MLO14) thus far, and second, the moss *Physcomitrella* also possesses clade I MLOs but does not produce true roots. In turn, the phylogenetic tree shows that the *Physcomitrella* MLO proteins are grouped in a clade I subclass, whereas *Selaginella*, a spikemoss (lycopod), harbors at least one MLO that groups to the MLOs from the seed-forming plants (fig. 1). In contrast to bryophytes, Lycopodiophyta are vascular plants that also form true roots. The emerging clade I phylogeny may therefore hint to at least a dual role of clade I MLO proteins, one being the regulation of root thigmomorphogenesis and the second being related to (an) unknown aspect(s) of moss and non-seed plant physiology.

Our phylogenetic data further suggest that clades II and IV appeared next in plant evolution, since the gymnosperms *P. abies* and *P. taeda* possess several MLOs from clade IV, and *P. abies* also has one clade II MLO. We hypothesize that these clades could be associated to the evolution of the seed, which is the embryonic plant protected by a coat and an evolutionary innovation associated with gymnosperms. However, it remains unclear which of these two clades is more ancient. Possibly, the addition of true fern (Monilophyta) MLOs to phylogenetic studies could help to unveil the exact relationship of these two clades. The presence of clade II and/or clade IV MLOs in ferns (which are non-seed plants) would argue against an association of these clades with seed functions. The genomes of the ferns *Azolla* and *Ceratopteris* are currently in progress and await finishing and release to the public (Sessa et al. 2014). Therefore, it should soon be possible to shed more light on the evolution of MLO clades II and IV.

Clades III, V, and VI appear to have evolved early in the angiosperm lineage, as the basal angiosperm *A. trichopoda* harbors MLOs from clades I–VI. Amborella is thought to have split from all other angiosperms early in angiosperm evolution and is considered a sister group to both monocots and eudicots (Drew et al. 2014). This finding suggests that MLOs of clades III, V, and VI are associated with the evolution of flowering plants. However, we did not find any monocot MLO protein that belongs to clade V or VI (fig. 1 and table 1), which we interpret as an indication of secondary loss of these clades in the monocot lineage. It should, however, be noted that only grasses (Poaceae) were included as monocot representatives in our study; non-grass monocots such as palms or orchids were not part of the dataset. Therefore, the proposed loss of MLO clades V and VI could be restricted to grasses. Inclusion of MLO sequences from a basal monocot species, for example, from the order Acorales or Alismatales, would also be useful to define better when monocots lost these clades. The function of clade IV MLOs seems to partially overlap with that of clade V MLOs. This is substantiated by the fact that mutants that are either defective in barley Mlo (*Hv*Mlo) (Büschges et al. 1997) or the three wheat *Ta*Mlo1 homeologs (Wang et al. 2014) from clade IV in monocots or Arabidopsis *At*MLO2 from clade V (Consonni et al. 2006) share similar phenotypes. *Hvmlo*, *Tamlo1A Tamlo1B Tamlo1D*, and *Atmlo2* mutant plants all show broad-spectrum powdery mildew resistance, and at least *Hvmlo* and *Atmlo2* further display early leaf senescence and spontaneous callose deposition. Intriguingly, a recent study revealed that transgenically expressed *Hv*Mlo restores powdery mildew susceptibility in the resistant tomato mutant *Slmlo1* (syn. *Slmlo6*), revealing the competence of clades IV and V MLOs for complementation across these clades (Appiano et al. 2015). This scenario differs from the situation in roots, where *At*MLO2 (a clade V MLO) expression failed to complement the aberrant thigmomorphogenesis phenotype of an *Atmlo4* (clade I MLO) mutant (Chen et al. 2009). Notably, many eudicot species seem to have lost clade IV MLOs during evolution. This may indicate that clades V and VI gradually took over the function of clade IV MLO proteins in the course of evolution.

Several lines of evidence point to a crucial role for clade III MLOs in flowering plants. For example, promoter–GUS reporter lines in Arabidopsis showed that most of the clade III *MLO*s were expressed in reproductive organs (Chen et al. 2006). In line with this, Arabidopsis *Atmlo7* mutants are deficient in proper pollen tube perception by the female gametophyte (Kessler et al. 2010). Furthermore, rice *Os*MLO12 is required for pollen hydration after contact of the pollen grain with a compatible stigma (Yi et al. 2014). The growth of pollen tubes towards the ovule through a stigma is a specific feature of angiosperms, as opposed to gymnosperms where the pollen grain (male gametophyte) germinates upon direct contact with the ovule. Therefore, we propose that clade III MLOs are key regulators of essential gametophytic functions during the fertilization process in angiosperms. The seventh clade was found to be present in eudicots only, including the basal

eudicots *N. nucifera* and *A. coerulea* (fig. 1 and table 1). Hence, MLOs from this clade might contribute to specific features of eudicot physiology. However, no mutant phenotype or function has ever been associated with a MLO protein from clade VII, rendering this hypothesis purely speculative.

Conserved MLO Motifs Are Associated with Regulation and TM Transport

An important aspect of this study is the identification of conserved motifs in MLO proteins. In our analysis, we substantially extended the set of identified conserved amino acids in the MLO protein (fig. 3; Elliott et al. 2005), allowing a far more detailed prediction of putative motifs and assigned functions. We observed the presence of several lysine/arginine-rich motifs in cytoplasmic domains 1, 2, and 4 (table 3 and supplementary fig. 6, Supplementary Material online). Lysine–arginine-rich motifs are commonly associated with nuclear import of cytosolic proteins (Görlich and Kutay 1999), which is not relevant for MLO proteins as they are integral membrane proteins. Additional processes where lysine and arginine have been described to be involved in are RNA binding (Bayer et al. 2005) and the interaction of membrane proteins with phosphatidic acid (Petersen et al. 2012). Phosphatidic acid binding seems a plausible function of these lysine–arginine motifs, since MLO proteins are membrane-bound and close interaction with phospholipids should be important for protein stability and integrity. Additional noteworthy motifs include RFR and TSF[GV][RK]RH in the second internal loop, FWF in the third cytoplasmic loop, and FG[IL][RK]SCF in the third apoplastic/luminal loop. Through GO association analysis against the *S. cerevisiae* proteome, these motifs were associated with TM transport of various potential molecules (table 3 and supplementary file 4, Supplementary Material online). This could indicate a function of MLO proteins in TM transport. However, these motifs are associated with different types of transporters. The FWF motif, for example, can be found in ion transporters, drug transporters or polyamine transporters in yeast, hence providing no lead toward putative substrates for potential MLO functions. Further, the motif ETXTLE in the second cytoplasmic region is associated with "regulation of biological processes" as GO term. In addition, we associated motifs that were found in the rather diverse domains (supplementary files 5 and 6, Supplementary Material online) with GO terms, where "TM transport" and "regulation" also appear as the most significant GO terms. This analysis brings forward the hypothesis that in general, MLO proteins may function either as transporters of organic or inorganic cargo and/or as regulators of molecular processes. It should be nevertheless considered that the functional assignment of the identified sequence motifs is based on data from baker's yeast, an organism that is very distantly related to plants and lacks any MLO proteins. To date, direct evidence for biochemical functions of MLO is not available.

MLO Is an Ancient and Diversified Eukaryotic Protein

We searched green algae and organisms from other kingdoms for MLO-like proteins. Indeed, we found MLO-like proteins in most groups of green algae, but also in several Chromalveolata, an Apusozoa and an Amoebozoa species (fig. 4, supplementary file 8, Supplementary Material online). On the other hand, Ophistokonta (including fungi and animals), Excavata (a group of unicellular eukaryotes), and Prokaryota species seem to lack MLO-like proteins. Since no definite MLO-like protein was identified in any prokaryote, the current data suggest that MLO is an ancient eukaryotic protein that diversified greatly in some unicellular eukaryotic lineages, but retained a specific pattern with higher levels of conservation (both sequence- and topology-wise) in land plants. These findings can be best explained with a scenario where MLO evolved in a photosynthetically active unicellular eukaryote that was an ancestor to both the Archaeplastida and Chromalveolata, as these lineages either have primary (Archaeplastida) or secondary (Chromalveolata) chloroplasts. However, the situation may be more complex than this. Several green algae, including the Charophycae and Charales (the closest relatives to land plants), do not possess any *MLO* genes according to BLAST searches. Furthermore, MLO-like proteins were found in some Chromalveolata groups, such as oomycetes and Apicomplexa, but species in many clades of this kingdom seem to be devoid of any MLO proteins. Finally, one amoeba and a protozoan species were found to possess an MLO-like protein, even though these organisms are more closely related to animals and fungi than to Archaeplastida and Chromalveolata. These seemingly contradictory findings may be explained in several ways. First, the number of available genomes is limited for many of the analyzed groups (e.g., only one glaucophyte genome was released yet, that of *Cyanophora biloba*). Second, some genomes are incomplete and/or highly fragmented (e.g., the genome assemblies of *V. carteri* and *C. biloba* comprise 11,403 and 60,119 contigs, respectively; Prochnik et al. 2010; Price et al. 2012), which could result in false negative BLAST results in some instances. Third, it is possible that our findings reflect true evolutionary patterns. These could be rationalized by secondary gene losses, convergent evolution of MLO-like proteins and/or horizontal gene transfer. As more and better annotated genomes from all kingdoms of life will become available, the phylogeny of MLO-like proteins should become clearer.

MLO Rosetta Stone Proteins—a Novel Lead to Decipher MLO Function?

Another intriguing finding of our study is the identification of putative MLO fusion proteins. For example, in CDART analysis we discovered six and based on InterPro domains 17 occurrences of presumptive MLO fusions with EF hands in non-land plant species (supplementary files 8 and 10, Supplementary

Material online, and fig. 6). EF hands are the basic building blocks of calmodulin, a Ca^{2+} binding protein that has been shown to physically interact with barley Mlo and to be required for barley Mlo function (Kim et al. 2002a, 2002b). It is well known that in some instances pairs of interacting proteins have counterparts in other organisms where both proteins are fused into a single polypeptide chain. Such cases are termed Rosetta stone sequences, in analogy to the original Egypt Rosetta stone, since they decipher the physical interaction of the respective protein pair (Marcotte et al. 1999). The fusion of two interacting proteins into one entity can have the advantage of increased interaction efficiency and specificity, since the corresponding partners are already in close spatial proximity.

Rosetta stone proteins may assist the assignment of functions to yet uncharacterized proteins (Date 2008). For example, the mammalian protein Nit1 was found to act in the same tumor suppressor pathway as Fhit, discovered through the invertebrate Rosetta stone protein NitFhit (Semba et al. 2006). The discovery of Rosetta stone-type MLO proteins could thus provide a lead for the identification of novel MLO interaction partners, and thus ultimately MLO function. In that sense, the MLO-EF hand fusion can be regarded as a "proof of concept". The occurrence of multiple MLO-EF hand fusion proteins in rather unrelated non-land plant lineages (Alveolata, Stamenopiles, Chlorophyta, and Apusozoa) suggests that this fusion happened multiple times independently during evolution. At least in some instances, these fusions rely on continuous (intron-free) open reading frames (ORFs; e.g., GenBank accessions ETL41083 and CCA21210), largely excluding the possibility of faulty gene annotation in these cases. Given the seeming advantages of Rosetta stone proteins outlined above, one may wonder why no MLO-calmodulin fusions are present in embryophytes. A possible explanation is that the separation of the two entities enables one calmodulin protein to serve multiple MLO proteins (note that MLO-like proteins are single copy proteins in most non-land plant species and present as medium-sized protein families in embryophytes). In addition or alternatively, different calmodulin isoforms might be engaged in the interaction with MLOs in land plants, which would potentially bring in a new level of fine-tuning MLO activities.

Besides the MLO-EF hand fusion, we detected few additional cases of MLO fusions, for example with glycosyl hydrolases, RNase H-like proteins, and zinc finger-domain proteins (fig. 6 and supplementary file 10, Supplementary Material online). This observation suggests that MLO proteins might interact with these proteins, thus participating in biological processes including these activities. However, most of these cases are based on single or few hits in the database. We can thus not rule out the possibility that these are false positives due to sequencing or annotation errors, for example, by joining adjacent ORFs *via* inadequate intron positioning. Transcript or proteome data will be required to convincingly substantiate the existence of these predicted protein fusions.

Concluding Remarks

Through the largely increased dataset of land plant MLOs provided in this study we were able to strengthen the support for seven phylogenetic clades in the embryophyte MLO protein family. Our analyses suggest that clade I is the most ancient among these MLO clades, and that clade III can be linked to the development of the fertilization process in flowering plants. We identified evolutionary patterns, namely conserved as well as clade-specific protein motifs, which we attributed to the putative MLO functions "TM transport" and "regulation". We were able to trace back MLO proteins to unicellular plastid-carrying eukaryotes, as many algae and chromalveolates possess MLO-like proteins, which are highly diversified with respect to amino acid sequence and predicted membrane topology. Finally, we identified predicted Rosetta stone-type MLO proteins that may provide a lead for forthcoming functional studies. Future challenges in the field of MLO research will include (a) further elucidation of the evolution of MLO proteins, including, for example, MLOs from ferns, (b) the identification of clade-specific processes of MLO proteins, especially regarding clades II, VI, and VII, (c) the verification or falsification of MLO function as regulator or transporter, and (d) to validate or reject the candidate interacting proteins emerging from predicted Rosetta stone MLO proteins.

Methods

Identification of MLO Proteins

Using Arabidopsis, tomato and rice MLO protein sequences representing all clades as a query, translated plant genomes were examined for MLO proteins using BLASTP (Altschul et al. 1997). The following databases were used for these searches: NCBI http://blast.ncbi.nlm.nih.gov/Blast.cgi (Altschul et al. 1997), http://solgenomics.net/ (*N. benthamiana* Genome v. 1.0.1), Phytozome v10.2 http://phytozome.jgi.doe.gov/pz/portal.html (*A. coerulea* v1.1, *P. patens* v3.0, *S. moellendorfii* v.1.0, *S. italica* v2.1), http://congenie.org/ (*P. abies* v1.0, *P. taeda* v1.0), http://www.amborella.org/ (Amborella genome scaffold v1.0), http://webblast.ipk-gatersleben.de/barley/ (*H. vulgare* v1). Database quests for non-land plant MLOs were performed in autumn 2015. Sequences were examined manually for apparent completeness and correctness against an alignment of published MLO protein sequences as reference, using BioEdit v7.1.11 (Hall 1999).

Alignments and Phylogenetic Analysis

For further analysis, all MLO amino acid sequences were aligned using ClustalW implemented in the MEGA v6.0 software (Tamura et al. 2013). The alignment was further optimized by manual inspection and curation. The phylogenetic

trees were generated with MEGA v6.0, using the Maximum-Likelihood, Maximum parsimony, Neighbor-Joining, and UPGMA methods based on the Jones–Taylor–Thornton (JTT) matrix model, with 1,000 bootstrap replications each. The consensus sequences were generated using BioEdit v.7.1.11 (Hall 1999). Minimal consensus sequences were generated with a threshold of 0%, local conservation levels were determined with consensus sequences using thresholds of 90% (high conservation) and 95% (quasi-invariant residues).

Motif Search and GO Analysis

Amino acid sequence motifs were analyzed as WebLogos, generated in the sequence logo generator (http://weblogo.berkeley.edu/, Crooks et al. 2004). Further, motif searches were conducted on amino acid sequence alignments using the MEME tool from the MEME suite (http://meme-suite.org/, Bailey and Elkan 1994; Bailey et al. 2015). Motifs identified by this procedure (supplementary files 4–6, Supplementary Material online) were subjected to analysis *via* the FIMO search tool from the MEME suite against the Ensembl Genomes database for *S. cerevisiae* (http://meme-suite.org/tools/fimo, Grant et al. 2011). The resulting gene list was analyzed for over-representation of GO terms by Generic GO Term Finder (http://go.princeton.edu/cgi-bin/GOTermFinder, Boyle et al. 2004) and annotated using the *S. cerevisiae* assembly April 2011 (sacCer3) UCSC Table Browser (http://genome.ucsc.edu/cgi-bin/hgTables, Karolchik et al. 2004).

Domain Architecture Searches

Using a clade I MLO protein sequence as the query (*At*MLO4), we searched the databases CDART (http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi, Geer et al. 2002) and InterPro (http://www.ebi.ac.uk/interpro/, Jones et al. 2014) for MLO domain architectures. MLO domain architecture reconnaissance was performed in autumn 2015.

Protein Topology Prediction

The topology of the nonland plant MLOs was analyzed *via* TOPCONS 2.0 (http://topcons.cbr.su.se/pred/, Tsirigos et al. 2015), allowing integrated (consensus) prediction of TM domains and signal peptides on the basis of several algorithms. Land plant MLOs were mined for ID protein regions by IUPred (http://iupred.enzim.hu/, Dosztányi et al. 2005).

Statistical Analysis

The statistics program R v3.1.2 (R foundation, https://www.r-project.org/, R Core Team 2014) was used for statistics, regression analysis and generation of boxplots, line graphs and pie charts. Based on the nature of our data, which shows non-normal distribution, unequal variance, and consists of small sample sizes, statistical analysis were performed *via* Generalized Linear Modeling (GLM; Crawley 2015). Means, medians, and standard deviations (SD) were calculated in R. The respective R scripts used in this study are given in supplementary file 14, Supplementary Material online.

Supplementary Material

Supplementary files S1–S14, supplementary tables S1–S3, and supplementary figures S1–S6 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This study was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft; DFG) [grant PA861/11-1 to R.P.].

Literature Cited

Acevedo-Garcia J, Kusch S, Panstruga R. 2014. Magical mystery tour: MLO proteins in plant immunity and beyond. New Phytol. 204:273–281.

Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Amborella Genome Project. 2013. The *Amborella* genome and the evolution of flowering plants. Science 342:1241089.

Appiano M, et al. 2015. Monocot and dicot MLO powdery mildew susceptibility factors are functionally conserved in spite of the evolution of class-specific molecular features. BMC Plant Biol. 15:257.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815.

Bai Y, et al. 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. Int. Conf. Intell. Syst. Mol. Biol. 2:28–36.

Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME suite. Nucleic Acids Res. 43:W39–W49.

Banks JA, et al. 2011. The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. Science 332:960–963.

Bayer TS, Booth LN, Knudsen SM, Ellington AD. 2005. Arginine-rich motifs present multiple interfaces for specific binding by RNA. RNA 11:1848–1857.

Becker B, Marin B. 2009. Streptophyte algae and the origin of embryophytes. Ann. Bot. 103:999–1004.

Bennetzen JL, et al. 2012. Reference genome sequence of the model plant *Setaria*. Nat. Biotechnol. 30:555–561.

Berg JA, et al. 2015. A transposable element insertion in the susceptibility gene *CsaMLO8* results in hypocotyl resistance to powdery mildew in cucumber. BMC Plant Biol. 15:243.

Bidzinski P, et al. 2014. Physiological characterization and genetic modifiers of aberrant root thigmomorphogenesis in mutants of *Arabidopsis thaliana MILDEW LOCUS O* genes. Plant Cell Environ. 37:2738–2753.

Blanchard H, et al. 1997. Structure of a calpain Ca(2+)-binding domain reveals a novel EF-hand and Ca(2+)-induced conformational changes. Nat. Struct. Biol. 4:532–538.

Bombarely A, et al. 2012. A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. Mol. Plant Microbe Interact 25:1523–1530.

Boyle EI, et al. 2004. GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene

Ontology terms associated with a list of genes. Bioinformatics 20:3710–3715.
Brown CJ, Johnson AK, Dunker AK, Daughdrill GW. 2011. Evolution and disorder. Curr. Opin. Struct. Biol. 21:441–446.
Büschges R, et al. 1997. The barley *Mlo* gene: a novel control element of plant pathogen resistance. Cell 88:695–705.
Chen Z, et al. 2006. Expression analysis of the *AtMLO* gene family encoding plant-specific seven-transmembrane domain proteins. Plant Mol Biol. 60:583–597.
Chen Z, et al. 2009. Two seven-transmembrane domain MILDEW RESISTANCE LOCUS O proteins cofunction in *Arabidopsis* root thigmomorphogenesis. Plant Cell 21:1972–1991.
Chen Y, Wang Y, Zhang H. 2014. Genome-wide analysis of the *mildew resistance locus* (*MLO*) gene family in tomato (*Solanum lycopersicum* L.). Plant Omics J. 7:87–93.
Consonni C, et al. 2006. Conserved requirement for a plant host cell protein in powdery mildew pathogenesis. Nat. Genet. 38:716–720.
Crawley MJ. 2015. Statistics. An Introduction Using R. 2nd edn. Imperial College London, UK. John Wiley & Sons, Ltd.
Crooks GE, Hon G, Chandonia J, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res. 14:1188–1190.
Date SV. 2008. The Rosetta stone method. Methods Mol. Biol. 453:169–180.
Deshmukh R, Singh VK, Singh BD. 2014. Comparative phylogenetic analysis of genome-wide *Mlo* gene family members from *Glycine max* and *Arabidopsis thaliana.* Mol. Genet. Genomics 289:345–359.
Devoto A, et al. 1999. Topology, subcellular localization, and sequence diversity of the Mlo family in plants. J. Biol. Chem. 274:34993–35004.
Devoto A, et al. 2003. Molecular phylogeny and evolution of the plant-specific seven-transmembrane MLO family. J. Mol. Evol. 56:77–88.
Dosztányi Z, Csizmok V, Tompa P, Simon I. 2005. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21:3433–3434.
Drew BT, et al. 2014. Another look at the root of the angiosperms reveals a familiar tale. Syst. Biol. 63:368–382.
Dyson HJ, Wright PE. 2005. Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell. Biol. 6:197–208.
Elliott C, Zhou F, Spielmeyer W, Panstruga R, Schulze-Lefert P. 2002. Functional conservation of wheat and rice *Mlo* orthologs in defense modulation to the powdery mildew fungus. Mol. Plant Microbe Interact 15:1069–1077.
Elliott C, et al. 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, Torregrosa L, Panstruga R, Dry IB. 2008. Identification of grapevine *MLO* gene candidates involved in susceptibility to powdery mildew. Funct. Plant Biol. 35:1255.
Geer LY, Domrachev M, Lipman DJ, Bryant SH. 2002. CDART: protein homology by domain architecture. Genome Res. 12:1619–1623.
Gifford JL, Walsh MP, Vogel HJ. 2007. Structures and metal-ion-binding properties of the Ca^{2+}-binding helix-loop-helix EF-hand motifs. Biochem. J. 405:199–221.
Goff SA, et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92–100.
Görlich D, Kutay U. 1999. Transport between the cell nucleus and the cytoplasm. Annu. Rev. Cell. Dev. Biol. 15:607–660.
Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. Bioinformatics 27:1017–1018.
Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41:95–98.
Huang S, et al. 2009. The genome of the cucumber, *Cucumis sativus* L. Nat. Genet. 41:1275–1281.
Humphry ME, et al. 2010. A regulon conserved in monocot and dicot plants defines a functional module in antifungal plant immunity. Proc. Natl Acad. Sci. U. S. A. 107:21896–21901.
Humphry M, Reinstädler A, Ivanov S, Bisseling T, Panstruga R. 2011. Durable broad-spectrum powdery mildew resistance in pea *er1* plants is conferred by natural loss-of-function mutations in *PsMLO1*. Mol. Plant Pathol. 12:866–878.
Iovieno P, et al. 2015. Structure, evolution and functional inference on the *Mildew Locus O* (*MLO*) gene family in three cultivated *Cucurbitaceae* spp. BMC Genomics 16:1112.
International Brachypodium Initiative. 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763–768.
International Wheat Genome Sequencing Consortium (IWGSC). 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345:1251788.
Jaillon O, et al. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467.
Jiao Y, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. Nature 473:97–100.
Jiwan D, Roalson EH, Main D, Dhingra A. 2013. Antisense expression of peach mildew resistance locus O (*PpMlo1*) gene confers cross-species resistance to powdery mildew in *Fragaria x ananassa.* Transgenic Res. 22:1119–1131.
Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240.
Jørgensen JH. 1992. Discovery, characterization and exploitation of Mlo powdery mildew resistance in barley. Euphytica 63:141–152.
Karolchik D, et al. 2004. The UCSC Table Browser data retrieval tool. Nucleic Acids Res. 32:D493–D496.
Kessler SA, et al. 2010. Conserved molecular components for pollen tube reception and fungal invasion. Science 330:968–971.
Kim MC, et al. 2002a. Calmodulin interacts with MLO protein to regulate defence against mildew in barley. Nature 416:447–451.
Kim MC, et al. 2002b. Mlo, a modulator of plant defense and cell death, is a novel calmodulin-binding protein. Isolation and characterization of a rice Mlo homologue. J. Biol. Chem. 277:19304–19314.
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.
Lalonde S, et al. 2010. A membrane protein/signaling protein interaction network for *Arabidopsis* version AMPv2. Front Physiol. 1:24.
Lamesch P, et al. 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. 40:D1202–D1210.
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, Griebel T, Jones AM, Kuhn H, Panstruga R. 2013. The role of *Arabidopsis* heterotrimeric G-protein subunits in MLO2 function and MAMP-triggered immunity. Mol. Plant Microbe Interact 26:991–1003.
Marcotte EM, et al. 1999. Detecting protein function and protein-protein interactions from genome sequences. Science 285:751–753.
Mayer KFX, et al. 2012. A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711–716.
Ming R, et al. 2013. Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.). Genome Biol. 14:R41.
Neale DB, et al. 2014. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. Genome Biol. 15:R59.

Nilsson J, Grahn M, Wright APH. 2011. Proteome-wide evidence for enhanced positive Darwinian selection within intrinsically disordered regions in proteins. Genome Biol. 12:R65.
Nystedt B, et al. 2013. The Norway spruce genome sequence and conifer genome evolution. Nature 497:579–584.
Panstruga R. 2005. Serpentine plant MLO proteins as entry portals for powdery mildew fungi. Biochem. Soc. Trans. 33:389–392.
Pessina S, et al. 2014. Characterization of the *MLO* gene family in Rosaceae and gene expression analysis in *Malus domestica*. BMC Genomics 15:618.
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.
Petersen J, et al. 2012. The lysine-rich motif of intrinsically disordered stress protein CDeT11-24 from *Craterostigma plantagineum* is responsible for phosphatidic acid binding and protection of enzymes from damaging effects caused by desiccation. J. Exp. Bot. 63:4919–4929.
Price DC, et al. 2012. *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. Science 335:843–847.
Prochnik SE, et al. 2010. Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. Science 329:223–226.
R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
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.
Rensing SA, et al. 2008. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. Science 319:64–69.
Schmutz J, et al. 2010. Genome sequence of the palaeopolyploid soybean. Nature 463:178–183.
Schnable PS, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115.
Semba S, et al. 2006. Biological functions of mammalian Nit1, the counterpart of the invertebrate NitFhit Rosetta stone protein, a possible tumor suppressor. J. Biol. Chem. 281:28244–28253.
Sessa EB, et al. 2014. Between two fern genomes. GigaScience 3:15.
Shulaev V, et al. 2011. The genome of woodland strawberry (*Fragaria vesca*). Nat. Genet. 43:109–116.
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30:2725–2729.
Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635–641.
Tsirigos KD, Peters C, Shu N, Käll L, Elofsson A. 2015. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. Nucleic Acids Res. 43:W401–W407.
Velasco R, et al. 2010. The genome of the domesticated apple (*Malus* × *domestica* Borkh.). Nat. Genet. 42:833–839.
Verde I, et al. 2013. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat. Genet. 45:487–494.
Wang Y, et al. 2014. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. Nat. Biotechnol 32:947–951.
Wolter M, Hollricher K, Salamini F, Schulze-Lefert P. 1993. The *mlo* resistance alleles to powdery mildew infection in barley trigger a developmentally controlled defence mimic phenotype. Mol. Gen. Genetics 239:122–128.
Yi J, An S, An G. 2014. *OsMLO12*, encoding seven transmembrane proteins, is involved with pollen hydration in rice. Plant Reprod 27:169–180.
Zhou SJ, Jing Z, Shi JL. 2013. Genome-wide identification, characterization, and expression analysis of the *MLO* gene family in *Cucumis sativus*. Genet. Mol. Res. 12:6565–6578.

Associate editor: Geoff McFadden