Lipid exchanges drove the evolution of mutualism during plant terrestrialization

Mélanie K Rich, Nicolas Vigneron, Cyril Libourel, Jean Keller, Xue Li, Mohsen Hajheidhari, Guru V Radhakrishnan, Aurélie Le Ru, Seydina Issa Diop, Giacomo Potente, et al.

Title: Lipid exchanges drove the evolution of mutualism during plant terrestrialization

Mélanie K. Rich[1†], Nicolas Vigneron[1†], Cyril Libourel[1], Jean Keller[1], Li Xue[2], Mohsen Hajheidhari[2], Guru V. Radhakrishnan[3], Aurélie Le Ru[4], Issa S Diop[5,6], Giacomo Potente[5,6], Elena Conti[5,6], Danny Duijsings[7], Aurélie Batut[8], Pauline Le Faouder[8], Kyoichi Kodama[9], Junko Kyozuka[9], Erika Sallet[10], Guillaume Bécard[1], Marta Rodrigez-Franco[11], Thomas Ott[11,12], Justine Bertrand-Michel[8], Giles ED Oldroyd[3,13], Péter Szövényi[5,6], Marcel Bucher[2], Pierre-Marc Delaux[1]*

[†]These authors contributed equally to the work.

[1]LRSV, Université de Toulouse, CNRS, UPS, Castanet-Tolosan, France.

[2]Institute for Plant Sciences, Cologne Biocenter, Cluster of Excellence on Plant Sciences, University of Cologne, D-50674 Cologne, Germany

[3]John Innes Centre, Norwich, UK.

[4]Fédération de Recherche 3450, Plateforme Imagerie, Pôle de Biotechnologie Végétale, Castanet-Tolosan, 31326, France

[5]Department of Systematic and Evolutionary Botany, University of Zurich, Zurich, Switzerland

[6]Zurich-Basel Plant Science Center, Zurich, Switzerland

[7]BaseClear BV, Leiden, The Netherlands

[8]MetaToulLipidomics Facility, INSERM UMR1048, Toulouse, France

[9]Graduate School of Life Sciences, Tohoku University, Sendai 980-8577, Japan.

[10]LIPM, Université de Toulouse, INRA, CNRS, Castanet-Tolosan, France.

[11]Cell Biology, Faculty of Biology, University of Freiburg, 79104 Freiburg, Germany.

[12]CIBSS – Centre for Integrative Biological Signalling Studies, University of Freiburg, 79104 Freiburg, Germany

[13]Crop Science Centre, University of Cambridge, Cambridge, UK.

* Corresponding author: pierre-marc.delaux@lrsv.ups-tlse.fr

Abstract:

Symbiosis with arbuscular mycorrhizal fungi (AMF) improves plant nutrition in most land plants and its contribution to the colonization of land by plants has been hypothesized. Here, we identified a conserved transcriptomic response to AMF among land plants, including the activation of lipid metabolism. Using gain of function, we show the transfer of lipids from the liverwort *Marchantia paleacea* to AMF and its direct regulation by the transcription factor WRINKLED (WRI). Arbuscules, the nutrient-exchange structures, were not formed in loss-of-function *wri* mutants in *M. paleacea* leading to aborted mutualism. Our results show the orthology of the symbiotic transfer of lipids across land plants and demonstrate that mutualism with arbuscular mycorrhizal fungi was present in the most recent ancestor of land plants 450 million years ago.

One Sentence Summary: By recruiting the general lipid metabolism the evolution of mutualism with fungi 450 million years ago allowed plants to colonize land.

Main text:

The stunning diversity of extant terrestrial ecosystems results from the diversification of plants following their colonization of emerged land 450 million years ago (*1*). This unique terrestrialization event required the aquatic algal ancestor of land plants to adapt to desiccation, direct UV radiation and a nutrient-poor habitat (*2*). Although they remain elusive, the evolutionary innovations that allowed terrestrialization to occur are predicted to be shared across the two main clades of land plants, the non-vascular Bryophytes and the vascular plants, and absent from their close algal relatives (*2*).

Most land plants benefit from associations with the mutualistic arbuscular mycorrhizal fungi (AMF), which facilitate nutrient and water uptake (*3*). Fossil records indicate that spores morphologically similar to those of extant AMF were present when terrestrialization occurred (*4*). The earliest plant macrofossils, which belong to an extinct plant lineage that preceded vascular plants, harbor intracellular structures morphologically similar to extant symbiotic structures (*5*, *6*). In addition, plant genes known to be essential for the symbiosis with AMF in angiosperms have co-evolved with the ability to associate with these fungi across land plants (*7*, *8*). These data support the original hypothesis that the association with AMF might represent one of the traits facilitating terrestrialization (*9*). However, definitive evidence of the ancestral character of the symbiosis with

AMF in the most recent common ancestor of the land plants, and the molecular mechanisms at its origin, remain to be discovered.

If symbiosis with AMF indeed predated the divergence between Bryophytes and vascular plants, we hypothesized that there would be a shared transcriptional response to AMF in those two clades. To test this, we generated RNA sequencing data of the Bryophyte *Marchantia paleacea* collected three, five and eight weeks after either inoculation with the AMF *Rhizophagus irregularis* or mock treatment (Fig. 1, Table S1). Reads were mapped on a *M. paleacea* genome assembly generated here by augmenting a previous Illumina-based assembly (*8*) with long read sequencing (Table S2). Increasing transcriptional response was observed along the time course of the interaction leading to more than 1200 differentially regulated genes with more than 1000 upregulated genes at eight weeks post-inoculation (Fig. 1, Table S3). We cross-referenced these results with transcriptomic datasets available for five phylogenetically diverse angiosperms (*10–14*). GO term enrichment analysis revealed a set of conserved processes upregulated during the symbiosis across land plants including fatty acid biosynthesis, transmembrane transport, oxidoreductase activity and proteolysis (Fig. 1A, Table S1). To further investigate the conservation of the response, we computed groups of orthologous genes across land plants and identified, for each orthogroup, whether one or more genes were upregulated during the symbiosis in each species (Fig. S1, Table S3). We found 47 orthogroups corresponding to 69 *M. paleacea* genes upregulated during the symbiosis with AMF both in this liverwort and in at least three of the five investigated angiosperms (Fig. 1B). These orthogroups contain genes with functions in diverse aspects of the symbiosis in angiosperms such as strigolactone biosynthesis, infection and lipid exchanges (Fig. 1B, Table S3). Such conservation of cellular reprogramming shows that the induction of a core genetic program during AM symbiosis (AMS) is conserved between land plant lineages spanning 450 million years of diversification.

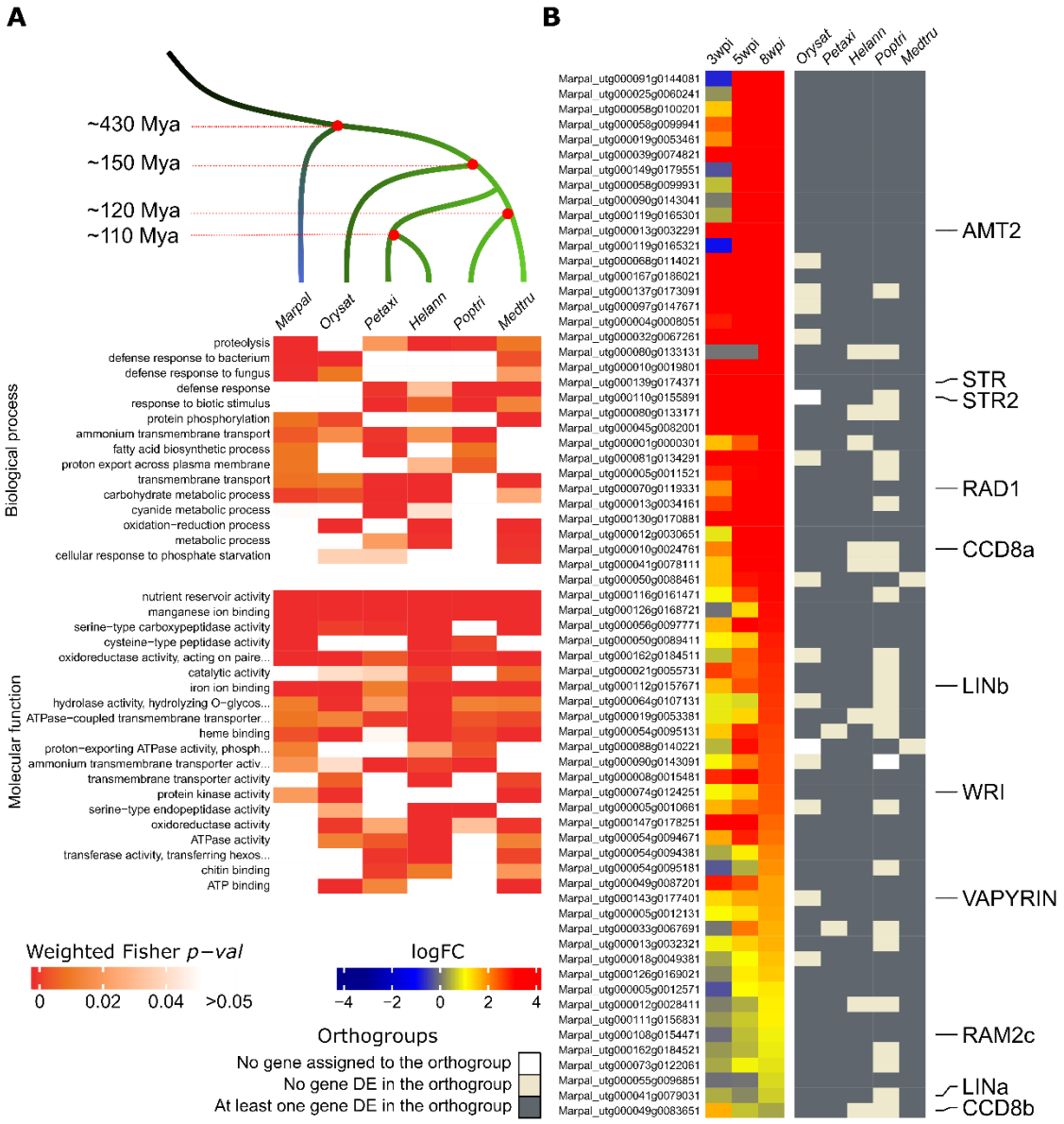

Figure 1. Conserved transcriptomic response to arbuscular mycorrhizal fungi across land plants. (**A**) Comparison of Gene Ontology terms enrichment during AMS in different plant species. GO terms enriched in at least 2 species are shown. Colors of the heatmap are weighted Fisher p-values of the respective GO terms. Marpal: *Marchantia* paleacea; Orysat: *Oryza sativa*; Petaxi: *Petunia axilaris*; Hellann: *Helianthus annuus*; Poptri: *Populus trichocarpa*; Medtru: *Medicago truncatula*. (**B**) Heatmap of gene expression during AMS in *M. paleacea*. Up-regulation of genes from the same orthogroups as the *M. paleacea* genes in five angiosperm species is indicated. Genes shown in this graph are significantly upregulated during AMS in *M. paleacea* in at least one of the time points and belong to an orthogroup where at least 3 other species present with AMS-upregulated genes (FDR < 0.05). Heatmap is shown as log2 fold change compared to mock treated plants. wpi: week post inoculation.

In angiosperms, *Required for Arbuscular Mycorrhiza 2 (RAM2), Stunted Arbuscule* (*STR)* and *STR2* are expressed in arbuscule-containing cells where they are essential for the biosynthesis and transfer of lipids to the AMF partners (*14*, *17–19*), which are fatty acid auxotrophs (*20*, *21*). In addition to the upregulation of *MpaRAM2c* and *MpaSTR/STR2,* genes corresponding to the GO term "Fatty acid biosynthesis" (GO:006633) were enriched in the subset of AMS-upregulated genes across host plants (Fig. 2A). The prevalence of lipid genes upregulated in *M. paleacea* suggests that lipid transfer from the host plant to the AMF occurs in Bryophytes, as is the case in angiosperms. To test this hypothesis, we compared the level of triacylglycerol (TAG) in AMF or mock inoculated *M. paleacea* thalli. TAGs acylated with two or three palmitic acids (C16) were enriched in AMF-colonized *M. paleacea* thalli compared to mock-treated plants (Fig. S2). To determine whether this accumulation of TAGs was mirrored by the presence of lipids in the AMF intracellular structures, sections of AMF-colonized *M. paleacea* were stained with a lipid-specific dye (Nile Red). Staining in the arbuscules and intracellular fungal hyphae was observed, indicating the accumulation of lipids in the AMF during the association with *M. paleacea* (Fig. 2D). Staining in the tissue area where arbuscules are hosted during symbiosis was absent in non-colonized plants (Fig. S3). To test whether the lipids observed in the AMF originate from *M. paleacea,* the fatty-acid metabolism in the arbuscule-containing cells was rewired by the expression of *Umbellularia californica Fatty Acid Thioestherase B* (*UcFatB)* under an arbuscule containing cell-specific promoter (Fig. S4) to produce a rare C12 acyl group, lauric acid (C12, (*14*)). In case transfer of lipids occurs between *M. paleacea* and the colonizing AMF, TAGs acylated with C12 should accumulate in both the colonized *UcFatB* expressing plant tissues and in the associated newly produced AMF spores. Mass-spectrometry analyses of AMF-inoculated wild-type and *UcFatB* lines revealed the accumulation of C44:2, a TAG containing one C12 and two C16, in the *UcFatB* lines, a compound barely detectable in the wild-type samples (Fig. 2B). Analysis of the AMF spores and hyphae connected to the *UcFatB* lines showed a similar enrichment (Fig. 2C). This demonstrates that lipids biosynthesized in arbuscule-containing cells of *M. paleacea* are transferred to the AMF. Altogether, these results indicate that, as in vascular plants, the symbiotic transfer of lipids from the host plant to AMF occurs in Bryophytes.

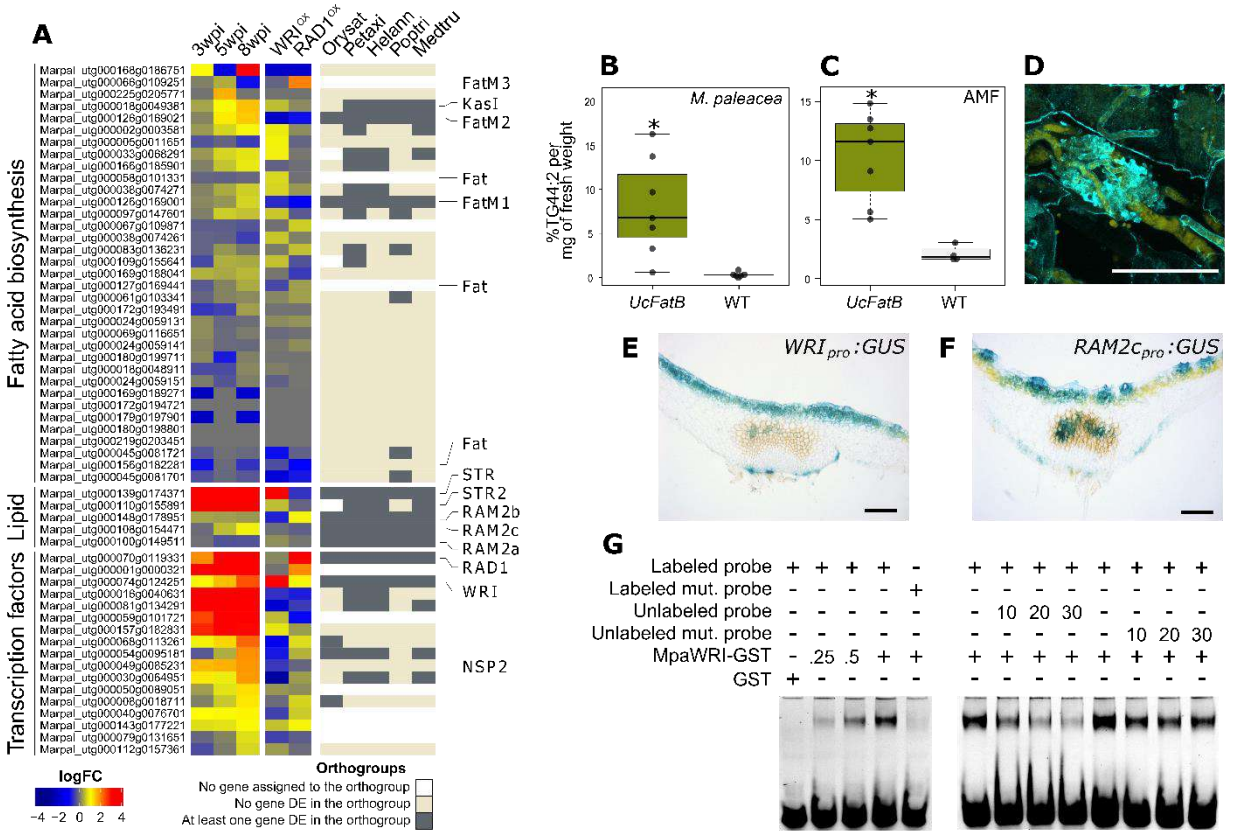

Fig. 2. WRI-induced symbiotic transfer of lipids is conserved across land plants. (**A**) Heatmap of log2 fold change of *Marchantia paleacea* genes in a time-course of AMF-inoculated compared to mock-treated plants, and in the *MpaWRI*^{OX} and *MpaRAD1*^{OX} lines. Are indicated genes from the fatty-acid biosynthesis GO term (006633), from the symbiotic lipid-transfer pathway, and *M. paleacea* transcription factors up-regulated in response to AMF. Up-regulation of genes from the same orthogroups as the *M. paleacea* genes in five angiosperm species is indicated. wpi: weeks post inoculation (**B, C**) Quantification of the C12 containing TAG 44:2 in samples from colonized WT and UcFatB *M. paleacea* lines (**B)** and the associated AMF (**C**). (**D**) Arbusculated cell from a section of *M. paleacea* stained with Nile red (yellow – neutral lipids) and calcofluor (blue – cell walls). Bar = 50 µm. (**E, F**) GUS reporter fusion of *MpaWRI* and *MpaRAM2c* promoters. Bar = 0.5 mm. (**G**) EMSA with GST-MpaWRI on a fragment of the *MpaRAM2c*$_{pro}$ (probe) and a mutated version (mut. probe).

To dissect the mechanisms underlying the symbiotic transfer of lipids, we mined the cross-referenced RNAseq data to identify potential transcriptional regulators of the symbiotic lipid pathway in arbuscule-containing cells (Fig. 2A). Two orthogroups contained transcriptional regulators upregulated in all six species colonized by AMF. These orthogroups contained genes known for their involvement in AMS in angiosperms: *Required for Arbuscular Development 1* (*RAD1)*, a GRAS transcriptional regulator (*22*) and the *WRINKLED* family of AP2 transcription factors (*23*, *24*). The *RAD1* orthogroup contained three *M. paleacea* genes and previous

phylogenies (*8*) had identified two of those genes (Marpal_utg000070g0119331, Marpal_utg000071g0120271) as putative RAD1 orthologs. From those two genes, only the latter (now renamed *MpaRAD1*) was upregulated during symbiosis (Fig. 2A, Table S3). Although members of the GRAS family are present in Zygnematophyceae, the closest algal relative to land plants, *RAD1* originates from a land plant-specific duplication and is absent from all the investigated algal genomes (Fig S5). Only one *WRINKLED* gene was found in *M. paleacea* (*MpaWRI,* Marpal_utg000074g0124251) and no *WRI* gene was found in algae (Fig. S6, Fig. S7, Fig. S8). Thus, the origin of *RAD1* and *WRI* coincides with terrestrialization. Fusion of *MpaRAD1* and *MpaWRI* promoters to a GUS reporter showed that both genes are expressed in arbuscule containing cells (Fig. S9). To determine whether MpaRAD1 or MpaWRI could regulate the symbiotic biosynthesis of fatty acids in *M. paleacea*, we generated lines overexpressing *MpaWRI* (*MpaWRI*^{OX}) and *MpaRAD1* (*MpaRAD1*^{OX}) and compared their transcriptome in the absence of AMF to empty-vector controls (Fig. 2A, Table S3). While in the *MpaRAD1*^{OX} lines no major gene expression reprogramming was observed, in *MpaWRI*^{OX} lines an enrichment of the GO term "Fatty acid biosynthesis" was observed among the differentially expressed genes (Fig. 2A, Table S1). This is reminiscent of the role of *WRINKLED* genes in controlling lipid biosynthesis in angiosperms, both in developmental and symbiotic contexts (*24*, *25*). Transient expression of *MpaWRI* in leaves of *Nicotiana benthamiana* led to accumulation of TAG, further confirming the link with the lipid metabolism (Fig. S10). In addition, the putative symbiosis-specific lipid transporter *STR*, was also upregulated in *MpaWRI*^{OX} lines in absence of AMF (Fig. 2A). Taking the reverse approach, we expressed a version of *MpaWRI* fused to the artificial repressor domain SRDX (*26*) (*MpaWRI*^{SRDX}). *MpaWRI*^{SRDX} lines showed developmental defects and a significant decrease in the expression of *MpaRAM2c,* the *RAM2* paralog from *M. paleacea* which is up-regulated during AMS (Fig. 2A, Fig. S11, Fig. S12) and whose expression pattern overlaps with that of *MpaWRI* (Fig. 2E-F). Within a 1 kb-portion of the *MpaRAM2c*$_{pro}$ that was sufficient for induction in arbuscules (Fig. S13), we identified a region containing known target sequences of angiosperm *WRINKLED* promoters (*27*). Using EMSA a 25 bp-long DNA sequence within this region was found to be sufficient for the direct binding of MpaWRI on *MpaRAM2c*$_{pro}$ (Fig. 2G, Fig. S13). Altogether, these data indicate that MpaWRI directly regulates the expression of the lipid metabolism in arbuscule-containing cells in *M. paleacea* leading to the transfer of fatty acids to the AMF.

To determine whether the function of *MpaWRI* is required for AMS, we generated a series of CRISPR/Cas9 *wri* mutants in *M. paleacea*. Five weeks after inoculation with the AMF *R. irregularis,* all inoculated empty vector control thalli were fully colonized and showed high density of arbuscules (Fig. 3A). By contrast, in five *wri* null mutants (*wri-1* to *5*) no fully developed arbuscules were found and AMF colonization structures were restricted to intraradical hyphae (Fig. 3). Similarly, absence of arbuscules was observed in the *MpaWRI*^{SRDX} lines (Fig. S12) and in the single *wri* allele obtained in another *M. paleacea* subspecies (Fig. S14). Two additional CRISPR alleles, *wri-6* and *wri-7*, with edited alleles of *MpaWRI* that did not impair the open reading frame of the gene (Fig. S15) showed an almost wild-type phenotype with more than 50 and 75% of the thalli harboring arbuscules respectively (Fig. 3C).

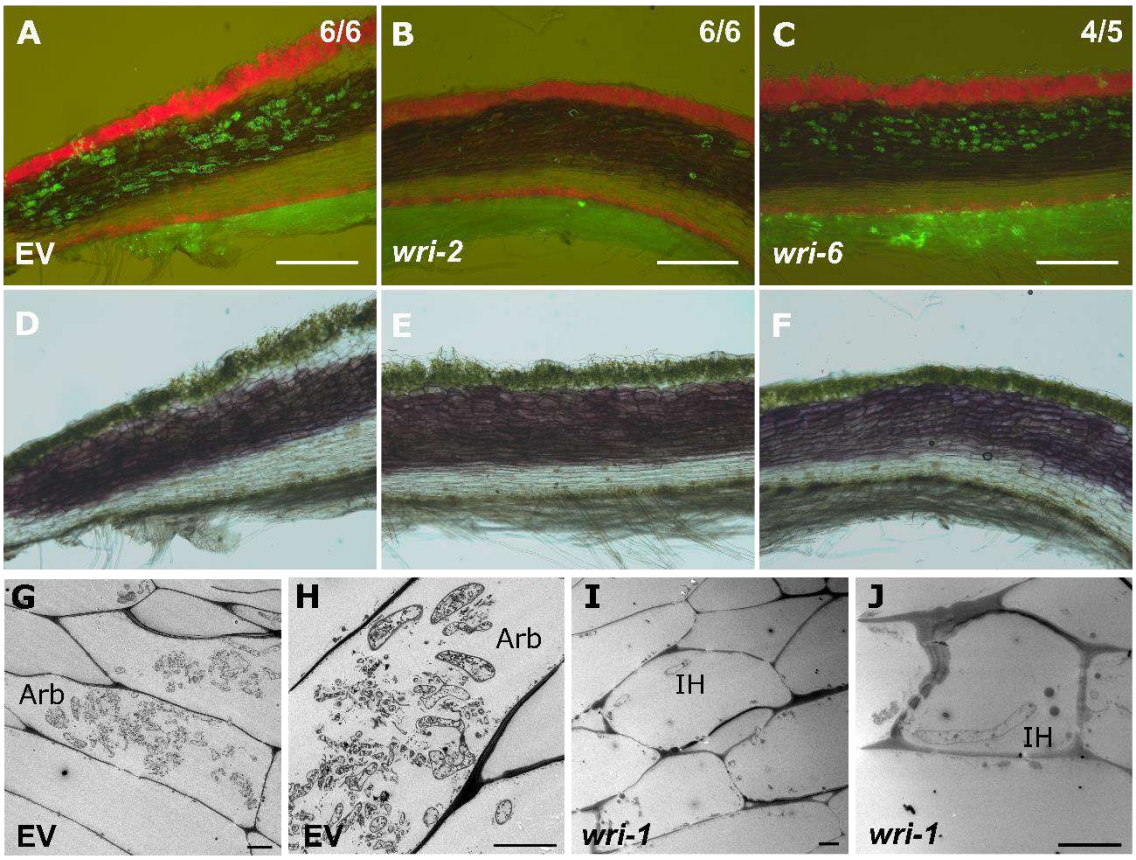

Fig. 3. WRI is necessary for AMS in *Marchantia paleacea*. (**A-F**) Transversal sections of empty vector control (**A, D**), *wri* null (**B, E**) or *wri* edited (**C, F**) *M. paleacea* thalli stained with WGA-FITC (top) or bright field (bottom). Proportion of thalli with the presented phenotype is indicated. Bar = 0.5 mm. (**G-I**) Transmission electron microscopy pictures of cells from empty vector control lines (**G, H**) or *wri-1* (**I, J**). In control plants, the central part of the thallus contains arbuscules (arb) while only intracellular hyphae (IH) are seen in the mutant. Bar = 10µm.

Our results demonstrate that the symbiotic regulation by *MpaWRI* of the biosynthesis of fatty-acid and their transfer to the symbiont are essential for mutualism with AMF in *M. paleacea*. In angiosperms, the *WRINKLED* orthogroup encompasses multiple paralogs (Fig. S6), including several that contribute to AMS (*23*, *24*). The less pronounced symbiotic defect observed in the angiosperm mutants compared to their liverwort counterpart is likely due to partial genetic redundancy. The fact that the same trait is regulated by orthologous genetic pathways in Bryophytes and in angiosperms indicates its conservation across land plants, covering 450 million years of diversification and dispersion. These results validate the long-standing proposed hypothesis that the mutualistic arbuscular mycorrhizal symbiosis was one of the traits that evolved in the first land plants, enabling terrestrialization.

References and Notes:

1. D. J. Beerling, *The Emerald Planet: How Plants Changed Earth's History* (Oxford University Press, 2008), vol. 83.

2. S. A. Rensing, How Plants Conquered Land. *Cell*. **181**, 964–966 (2020).

3. S. E. Smith, D. Read, in *Mycorrhizal Symbiosis (Third Edition)* (Academic Press, London, 2008), pp. 91–116.

4. D. Redecker, R. Kodner, L. E. Graham, Glomalean Fungi from the Ordovician. *Science*. **289**, 1920–1921 (2000).

5. W. Remy, T. N. Taylor, H. Hass, H. Kerp, Four hundred-million-year-old vesicular arbuscular mycorrhizae. *Proc Natl Acad Sci U S A*. **91**, 11841–11843 (1994).

6. P. Kenrick, P. R. Crane, The origin and early evolution of plants on land. *Nature*. **389**, 33–39 (1997).

7. P.-M. Delaux, G. V. Radhakrishnan, D. Jayaraman, J. Cheema, M. Malbreil, J. D. Volkening, H. Sekimoto, T. Nishiyama, M. Melkonian, L. Pokorny, C. J. Rothfels, H. W. Sederoff, D. W. Stevenson, B. Surek, Y. Zhang, M. R. Sussman, C. Dunand, R. J. Morris, C. Roux, G. K.-S. Wong, G. E. D. Oldroyd, J.-M. Ané, Algal ancestor of land plants was preadapted for symbiosis. *Proc Natl Acad Sci U S A*. **112**, 13390–13395 (2015).

8. G. V. Radhakrishnan, J. Keller, M. K. Rich, T. Vernié, D. L. Mbadinga Mbadinga, N. Vigneron, L. Cottret, H. S. Clemente, C. Libourel, J. Cheema, A.-M. Linde, D. M. Eklund, S. Cheng, G. K. S. Wong, U. Lagercrantz, F.-W. Li, G. E. D. Oldroyd, P.-M. Delaux, An ancestral signalling pathway is conserved in intracellular symbioses-forming plant lineages. *Nature Plants*, 1–10 (2020).

9. K. A. Pirozynski, D. W. Malloch, The origin of land plants: A matter of mycotrophism. *Biosystems*. **6**, 153–164 (1975).

10. V. Fiorilli, M. Vallino, C. Biselli, A. Faccio, P. Bagnaresi, P. Bonfante, Host and non-host roots in rice: cellular and molecular approaches reveal differential responses to arbuscular mycorrhizal fungi. *Front Plant Sci*. **6** (2015), doi:10.3389/fpls.2015.00636.

11. M. K. Rich, P.-E. Courty, C. Roux, D. Reinhardt, Role of the GRAS transcription factor ATA/RAM1 in the transcriptional reprogramming of arbuscular mycorrhiza in Petunia hybrida. *BMC Genomics*. **18**, 589 (2017).

12. A. Vangelisti, A. Turrini, C. Sbrana, L. Avio, T. Giordani, L. Natali, M. Giovannetti, A. Cavallini, Gene expression in Rhizoglomus irregulare at two different time points of mycorrhiza establishment in Helianthus annuus roots, as revealed by RNA-seq analysis. *Mycorrhiza*. **30**, 373–387 (2020).

13. Transcriptome analysis of the Populus trichocarpa–Rhizophagus irregularis Mycorrhizal Symbiosis: Regulation of Plant and Fungal Transportomes under Nitrogen Starvation. *Plant Cell Physiol*. **58**,1003–1017. (2017)

14. L. H. Luginbuehl, G. N. Menard, S. Kurup, H. V. Erp, G. V. Radhakrishnan, A. Breakspear, G. E. D. Oldroyd, P. J. Eastmond, Fatty acids in arbuscular mycorrhizal fungi are synthesized by the host plant. *Science*, eaan0081 (2017).

15. J. L. Morris, M. N. Puttick, J. W. Clark, D. Edwards, P. Kenrick, S. Pressel, C. H. Wellman, Z. Yang, H. Schneider, P. C. J. Donoghue, The timescale of early land plant evolution. *PNAS*. **115**, E2274–E2283 (2018).

16. F. Lutzoni, M. D. Nowak, M. E. Alfaro, V. Reeb, J. Miadlikowska, M. Krug, A. E. Arnold, L. A. Lewis, D. L. Swofford, D. Hibbett, K. Hilu, T. Y. James, D. Quandt, S. Magallón, Contemporaneous radiations of fungi and plants linked to symbiosis. *Nature Communications*. **9**, 5451 (2018).

17. Y. Jiang, W. Wang, Q. Xie, N. Liu, L. Liu, D. Wang, X. Zhang, C. Yang, X. Chen, D. Tang, E. Wang, Plants transfer lipids to sustain colonization by mutualistic mycorrhizal and parasitic fungi. *Science*, eaam9970 (2017).

18. A. Bravo, M. Brands, V. Wewer, P. Dörmann, M. J. Harrison, Arbuscular mycorrhiza-specific enzymes FatM and RAM2 fine-tune lipid biosynthesis to promote development of arbuscular mycorrhiza. *New Phytol*. **214**, 1631–1645 (2017).

19. A. Keymer, P. Pimprikar, V. Wewer, C. Huber, M. Brands, S. L. Bucerius, P.-M. Delaux, V. Klingl, E. von Röpenack-Lahaye, T. L. Wang, W. Eisenreich, P. Dörmann, M. Parniske, C. Gutjahr, Lipid transfer from plants to arbuscular mycorrhiza fungi. *eLife*. **6**, e29107 (2017).

20. V. Wewer, M. Brands, P. Dörmann, Fatty acid synthesis and lipid metabolism in the obligate biotrophic fungus Rhizophagus irregularis during mycorrhization of Lotus japonicus. *Plant J*. **79**, 398–412 (2014).

21. M. M. C, M. Krüger, C. Krüger, Y. Wang, J. E. Stajich, J. Keller, E. C. H. Chen, G. Yildirir, M. Villeneuve-Laroche, C. Roux, P.-M. Delaux, N. Corradi, The genome of Geosiphon pyriformis reveals ancestral traits linked to the emergence of the arbuscular mycorrhizal symbiosis. *Current Biology*. **0** (2021), doi:10.1016/j.cub.2021.01.058.

22. L. Xue, H. Cui, B. Buer, V. Vijayakumar, P.-M. Delaux, S. Junkermann, M. Bucher, Network of GRAS Transcription Factors Involved in the Control of Arbuscule Development in Lotus japonicus. *Plant Physiol.* **167**, 854–871 (2015).

23. Y. Jiang, Q. Xie, W. Wang, J. Yang, X. Zhang, N. Yu, Y. Zhou, E. Wang, Medicago AP2-domain Transcription Factor WRI5a Is a Master Regulator of Lipid Biosynthesis and Transfer During Mycorrhizal Symbiosis. *Molecular Plant* (2018), doi:10.1016/j.molp.2018.09.006.

24. L. Xue, L. Klinnawee, Y. Zhou, G. Saridis, V. Vijayakumar, M. Brands, P. Dörmann, T. Gigolashvili, F. Turck, M. Bucher, AP2 transcription factor CBX1 with a specific function in

symbiotic exchange of nutrients in mycorrhizal Lotus japonicus. *PNAS*. **115**, E9239–E9246 (2018).

25. A. To, J. Joubès, G. Barthole, A. Lécureuil, A. Scagnelli, S. Jasinski, L. Lepiniec, S. Baud, WRINKLED Transcription Factors Orchestrate Tissue-Specific Regulation of Fatty Acid Biosynthesis in Arabidopsis[W]. *Plant Cell*. **24**, 5007–5023 (2012).

26. K. Hiratsu, K. Matsui, T. Koyama, M. Ohme-Takagi, Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in Arabidopsis. *Plant J.* **34**, 733–739 (2003).

27. N. Fukuda, Y. Ikawa, T. Aoyagi, A. Kozaki, Expression of the genes coding for plastidic acetyl-CoA carboxylase subunits is regulated by a location-sensitive transcription factor binding site. *Plant Mol Biol*. **82**, 473–483 (2013).

28. C. Balzergue, V. Puech-Pagès, G. Bécard, S. F. Rochange, The regulation of arbuscular mycorrhizal symbiosis by phosphate in pea involves early and systemic signalling events. *J Exp Bot*. **62**, 1049–1060 (2011).

29. R. Nagar, High purity, high molecular weight DNA extraction from rust spores via CTAB based DNA precipitation for long read sequencing (2018), doi:10.17504/protocols.io.n5ydg7w.

30. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. **34**, 3094–3100 (2018).

31. H. Li, Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics*. **32**, 2103–2110 (2016).

32. R. Vaser, I. Sović, N. Nagarajan, M. Šikić, Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).

33. B. J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young, A. M. Earl, Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE*. **9**, e112963 (2014).

34. Fast gapped-read alignment with Bowtie 2 | Nature Methods, (available at https://www.nature.com/articles/nmeth.1923).

35. BlobTools: Interrogation of genome assemblies | F1000Research, (available at https://f1000research.com/articles/6-1287).

36. S. Foissac, J. Gouzy, S. Rombauts, C. Mathe, J. Amselem, L. Sterck, Y. V. de Peer, P. R. and T. Schiex, Genome Annotation in Plants and Fungi: EuGene as a Model Platform. *Current Bioinformatics*. **3** (2008), pp. 87–97.

37. P. Lamesch, T. Z. Berardini, D. Li, D. Swarbreck, C. Wilks, R. Sasidharan, R. Muller, K. Dreher, D. L. Alexander, M. Garcia-Hernandez, A. S. Karthikeyan, C. H. Lee, W. D. Nelson, L. Ploetz, S. Singh, A. Wensel, E. Huala, The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40**, D1202–D1210 (2012).

38. Y. Pecrix, S. E. Staton, E. Sallet, C. Lelandais-Brière, S. Moreau, S. Carrère, T. Blein, M.-F. Jardinaud, D. Latrasse, M. Zouine, M. Zahm, J. Kreplak, B. Mayjonade, C. Satgé, M. Perez, S. Cauet, W. Marande, C. Chantry-Darmon, C. Lopez-Roques, O. Bouchez, A. Bérard, F. Debellé, S. Muños, A. Bendahmane, H. Bergès, A. Niebel, J. Buitink, F. Frugier, M. Benhamed, M. Crespi, J. Gouzy, P. Gamas, Whole-genome landscape of Medicago truncatula symbiotic genes. *Nature Plants*. **4**, 1017–1025 (2018).

39. W. Bao, K. K. Kojima, O. Kohany, Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA.* **6**, 11 (2015).

40. J. H. Leebens-Mack, M. S. Barker, E. J. Carpenter, M. K. Deyholos, M. A. Gitzendanner, S. W. Graham, I. Grosse, Z. Li, M. Melkonian, S. Mirarab, M. Porsch, M. Quint, S. A. Rensing, D. E. Soltis, P. S. Soltis, D. W. Stevenson, K. K. Ullrich, N. J. Wickett, L. DeGironimo, P. P. Edger, I. E. Jordon-Thaden, S. Joya, T. Liu, B. Melkonian, N. W. Miles, L. Pokorny, C. Quigley, P. Thomas, J. C. Villarreal, M. M. Augustin, M. D. Barrett, R. S. Baucom, D. J. Beerling, R. M. Benstein, E. Biffin, S. F. Brockington, D. O. Burge, J. N. Burris, K. P. Burris, V. Burtet-Sarramegna, A. L. Caicedo, S. B. Cannon, Z. Çebi, Y. Chang, C. Chater, J. M. Cheeseman, T. Chen, N. D. Clarke, H. Clayton, S. Covshoff, B. J. Crandall-Stotler, H. Cross, C. W. dePamphilis, J. P. Der, R. Determann, R. C. Dickson, V. S. Di Stilio, S. Ellis, E. Fast, N. Feja, K. J. Field, D. A. Filatov, P. M. Finnegan, S. K. Floyd, B. Fogliani, N. García, G. Gâteblé, G. T. Godden, F. (Qi Y. Goh, S. Greiner, A. Harkess, J. M. Heaney, K. E. Helliwell, K. Heyduk, J. M. Hibberd, R. G. J. Hodel, P. M. Hollingsworth, M. T. J. Johnson, R. Jost, B. Joyce, M. V. Kapralov, E. Kazamia, E. A. Kellogg, M. A. Koch, M. Von Konrat, K. Könyves, T. M. Kutchan, V. Lam, A. Larsson, A. R. Leitch, R. Lentz, F.-W. Li, A. J. Lowe, M. Ludwig, P. S. Manos, E. Mavrodiev, M. K. McCormick, M. McKain, T. McLellan, J. R. McNeal, R. E. Miller, M. N. Nelson, Y. Peng, P. Ralph, D. Real, C. W. Riggins, M. Ruhsam, R. F. Sage, A. K. Sakai, M. Scascitella, E. E. Schilling, E.-M. Schlösser, H. Sederoff, S. Servick, E. B. Sessa, A. J. Shaw, S. W. Shaw, E. M. Sigel, C. Skema, A. G. Smith, A. Smithson, C. N. Stewart, J. R. Stinchcombe, P. Szövényi, J. A. Tate, H. Tiebel, D. Trapnell, M. Villegente, C.-N. Wang, S. G. Weller, M. Wenzel, S. Weststrand, J. H. Westwood, D. F. Whigham, S. Wu, A. S. Wulff, Y. Yang, D. Zhu, C. Zhuang, J. Zuidof, M. W. Chase, J. C. Pires, C. J. Rothfels, J. Yu, C. Chen, L. Chen, S. Cheng, J. Li, R. Li, X. Li, H. Lu, Y. Ou, X. Sun, X. Tan, J. Tang, Z. Tian, F. Wang, J. Wang, X. Wei, X. Xu, Z. Yan, F. Yang, X. Zhong, F. Zhou, Y. Zhu, Y. Zhang, S. Ayyampalayam, T. J. Barkman, N. Nguyen, N. Matasci, D. R. Nelson, E. Sayyari, E. K. Wafula, R. L. Walls, T. Warnow, H. An, N. Arrigo, A. E. Baniaga, S. Galuska, S. A. Jorgensen, T. I. Kidder, H. Kong, P. Lu-Irving, H. E. Marx, X. Qi, C. R. Reardon, B. L. Sutherland, G. P. Tiley, S. R. Welles, R. Yu, S. Zhan, L. Gramzow, G. Theißen, G. K.-S. Wong, One Thousand Plant Transcriptomes Initiative, One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*. **574**, 679–685 (2019).

41. C. P. Humphreys, P. J. Franks, M. Rees, M. I. Bidartondo, J. R. Leake, D. J. Beerling, Mutualistic mycorrhiza-like symbiosis in the most ancient group of land plants. *Nature Communications*. **1**, 103 (2010).

42. M. Pertea, D. Kim, G. Pertea, J. T. Leek, S. L. Salzberg, Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie, and Ballgown. *Nat Protoc*. **11**, 1650–1667 (2016).

43. G. Pertea, M. Pertea, GFF Utilities: GffRead and GffCompare. *F1000Res*. **9** (2020), doi:10.12688/f1000research.23297.2.

44. C. Cabau, F. Escudié, A. Djari, Y. Guiguen, J. Bobe, C. Klopp, Compacting and correcting Trinity and Oases RNA-Seq de novo assemblies. *PeerJ*. **5**, e2988 (2017).

45. M. H. Schulz, D. R. Zerbino, M. Vingron, E. Birney, Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*. **28**, 1086–1092 (2012).

46. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. **37**, 907–915 (2019).

47. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**, 2078–2079 (2009).

48. S. Anders, P. T. Pyl, W. Huber, HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. **31**, 166–169 (2015).

49. J. T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, J. P. Mesirov, Integrative Genomics Viewer. *Nat Biotechnol*. **29**, 24–26 (2011).

50. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. **26**, 139–140 (2010).

51. D. J. McCarthy, Y. Chen, G. K. Smyth, Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. **40**, 4288–4297 (2012).

52. D. M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*. **20**, 238 (2019).

53. P. Jones, D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.-Y. Yong, R. Lopez, S. Hunter, InterProScan 5: genome-scale protein function classification. *Bioinformatics*. **30**, 1236–1240 (2014).

54. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: architecture and applications. *BMC Bioinformatics*. **10**, 421 (2009).

55. K. Katoh, D. M. Standley, MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol*. **30**, 772–780 (2013).

56. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. **25**, 1972–1973 (2009).

57. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol*. **32**, 268–274 (2015).

58. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*. **14**, 587–589 (2017).

59. S. Guindon, J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst Biol*. **59**, 307–321 (2010).

60. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol*. **35**, 518–522 (2018).

61. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*. **47**, W256–W259 (2019).

62. S. R. Eddy, Accelerated Profile HMM Searches. *PLOS Computational Biology*. **7**, e1002195 (2011).

63. M. Dipp-Álvarez, A. Cruz-Ramírez, A Phylogenetic Study of the ANT Family Points to a preANT Gene as the Ancestor of Basal and euANT Transcription Factors in Land Plants. *Front. Plant Sci.* **10** (2019), doi:10.3389/fpls.2019.00017.

64. S. B. Carey, J. Jenkins, J. T. Lovell, F. Maumus, A. Sreedasyam, A. C. Payton, S. Shu, G. P. Tiley, N. Fernandez-Pozo, K. Barry, C. Chen, M. Wang, A. Lipzen, C. Daum, C. A. Saski, J. C. McBreen, R. E. Conrad, L. M. Kollar, S. Olsson, S. Huttunen, J. B. Landis, J. G. Burleigh, N. J. Wickett, M. G. Johnson, S. A. Rensing, J. Grimwood, J. Schmutz, S. F. McDaniel, *bioRxiv*, in press, doi:10.1101/2020.07.03.163634.

65. J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G. A. Salazar, E. L. L. Sonnhammer, S. C. E. Tosatto, L. Paladin, S. Raj, L. J. Richardson, R. D. Finn, A. Bateman, Pfam: The protein families database in 2021. *Nucleic Acids Research*. **49**, D412–D419 (2021).

66. F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, D. G. Higgins, Fast, scalable generation of high-

quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*. **7**, 539 (2011).

67. N. J. Patron, D. Orzaez, S. Marillonnet, H. Warzecha, C. Matthewman, M. Youles, O. Raitskin, A. Leveau, G. Farré, C. Rogers, A. Smith, J. Hibberd, A. A. R. Webb, J. Locke, S. Schornack, J. Ajioka, D. C. Baulcombe, C. Zipfel, S. Kamoun, J. D. G. Jones, H. Kuhn, S. Robatzek, H. P. Van Esse, D. Sanders, G. Oldroyd, C. Martin, R. Field, S. O'Connor, S. Fox, B. Wulff, B. Miller, A. Breakspear, G. Radhakrishnan, P.-M. Delaux, D. Loqué, A. Granell, A. Tissier, P. Shih, T. P. Brutnell, W. P. Quick, H. Rischer, P. D. Fraser, A. Aharoni, C. Raines, P. F. South, J.-M. Ané, B. R. Hamberger, J. Langdale, J. Stougaard, H. Bouwmeester, M. Udvardi, J. A. H. Murray, V. Ntoukakis, P. Schäfer, K. Denby, K. J. Edwards, A. Osbourn, J. Haseloff, Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *New Phytol*. **208**, 13–19 (2015).

68. F. Althoff, S. Kopischke, O. Zobell, K. Ide, K. Ishizaki, T. Kohchi, S. Zachgo, Comparison of the MpEF1α and CaMV35 promoters for application in Marchantia polymorpha overexpression studies. *Transgenic Res*. **23**, 235–244 (2014).

69. J.-F. Li, J. E. Norville, J. Aach, M. McCormack, D. Zhang, J. Bush, G. M. Church, J. Sheen, Multiplex and homologous recombination–mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. *Nature Biotechnology*. **31**, 688–691 (2013).

70. K. Ono, K. Ohyama, O. L. Gamborg, Regeneration of the liverwort Marchantia polymorpha L. From protoplasts isolated from cell suspension culture. *Plant Science Letters*. **14**, 225–229 (1979).

71. F. Miart, T. Desprez, E. Biot, H. Morin, K. Belcram, H. Höfte, M. Gonneau, S. Vernhettes, Spatio-temporal analysis of cellulose synthesis during cell plate formation in Arabidopsis. *The Plant Journal*. **77**, 71–84 (2014).

72. E. G. Bligh, W. J. Dyer, A rapid method of total lipid extraction and purification. *Can J Biochem Physiol*. **37**, 911–917 (1959).

73. A. Barrans, X. Collet, R. Barbaras, B. Jaspard, J. Manent, C. Vieu, H. Chap, B. Perret, Hepatic lipase induces the formation of pre-beta 1 high density lipoprotein (HDL) from triacylglycerol-rich HDL2. A study comparing liver perfusion to in vitro incubation with lipases. *J. Biol. Chem.* **269**, 11572–11577 (1994).

74. J. M. Lillington, D. J. Trafford, H. L. Makin, A rapid and simple method for the esterification of fatty acids and steroid carboxylic acids prior to gas-liquid chromatography. *Clin Chim Acta*. **111**, 91–98 (1981).

75. Y.-C. Fang, A.-C. McGRAW, H. Modjo, J. W. Hendrix, A Procedure for Isolation of Single-Spore Cultures of Certain Endomycorrhizal Fungi. *New Phytologist*. **95**, 107–114 (1983).

Acknowledgements: We thank Tatiana Vernié and Sebastian Schornack for fruitful discussions and comments on the manuscript, and Ben Miller for providing the Cas9 Level-0 construct. We are grateful to the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, doi: 10.15454/1.5572369328961167E12) for providing computing resources and to the MetaToul-Lipidomique Core Facility (I2MC, Inserm 1048, Toulouse, France). **Funding:** This work was supported by the Bill & Melinda Gates Foundation and UK Foreign, Commonwealth and Development Office as the Engineering Nitrogen Symbiosis for Africa (OPP1172165) to G.E.D.O., P-M.D. and T.O., by the Agence Nationale de la Recherche (ANR) grant EVOLSYM (ANR-17-CE20-0006-01) and by the Université Paul Sabatier ATP2016 grant to P.-M.D, the MetaboHUB-ANR-11-INBS-0010 grant to J.B-M, by the URPP Evolution in Action of the University of Zurich (PSZ, ISD), funding from The Forschungskredit (University of Zurich, PSZ, ISD), grants of the Swiss National Science Foundation (PSZ 160004, 131726), the EU's Horizon 2020 Research and Innovation Programme (ISD, GP, PSZ, EC and DU, PlantHUB-No. 722338), and the Georges and Antoine Claraz Foundation (GP, ISD, PSZ). This study was supported by the German Research Foundation (DFG) under Germany's Excellence Strategy (CIBSS – EXC-2189 – Project ID 39093984). The TEM (funded by the DFG grant INST 39/1153-1) is operated by the University of Freiburg, Faculty of Biology, as a partner unit within the Microscopy and Image Analysis Platform (MIAP) and the Life Imaging Center (LIC), Freiburg. This research was supported by the German Science Foundation Grant BU-2250/12-1 (to M.B.) and by the German Ministry of Education and Research (BMBF, project no. 031B0200 A to M.B.) within the frame of RECONSTRUCT. The Laboratoire de Recherche en Sciences Végétales (LRSV) belongs to the TULIP Laboratoire d'Excellence (ANR-10-LABX-41). Author contributions: M.K.R, N.V. P-M.D., P.S., E.C., D.D., J.Ke., J.Ky., C.L., G.E.D.O., J.B-M., G.B., M.B., M.H. L.X., designed the experiments. M.K.R., N.V., G.V.R., C.L., J.Ke., K.K., I.S.D., E.S., M.R-F., A.L.R., P.L.F., A.B., M.H. L.X., G.P., P.S. conducted the experiments. M.K.R., N.V., P-M.D., C.L., J.Ke., T.O., J.B-M., P.L.F., A.B., M.B., M.H. L.X., analyzed the results. M.K.R, N.V. and P-M.D. wrote the manuscript. All co-authors edited and commented on the initial draft. **Competing interests:** The authors declare no competing interests. **Data availability:** Genome assembly, annotation file and gene models are available through a genome browser at Marpal_GenomeBrowser and on NCBI (Bioproject PRJNA362997). Raw Oxford Nanopore data was submitted under the NCBI SRA accession SRR13188886. Raw sequencing RNAseq data can be found under NCBI SRA accession numbers from SRR13077953 to SRR13077969 corresponding to the Bioproject PRJNA362997 (*Marchantia paleacea* genome). This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession MUAA00000000. The version described in this paper is version MUAA02000000. All other data are available in the main paper or the supplement.

Supplementary Materials:

Materials and Methods

Figures S1-S18

Tables S1-S8

References (*26-75*)

Science
AAAS

Supplementary Materials for

Lipid exchanges drove the evolution of mutualism during plant terrestrialisation

Mélanie K. Rich[1†], Nicolas Vigneron[1†], Cyril Libourel[1], Jean Keller[1], Li Xue[2], Mohsen Hajheidhari[2], Guru V. Radhakrishnan[3], Aurélie Le Ru[4], Issa S Diop[5,6], Giacomo Potente[5,6], Elena Conti[5,6], Danny Duijsings[7], Aurélie Batut[8], Pauline Le Faouder[8], Kyoichi Kodama[9], Junko Kyozuka[9], Erika Sallet[10], Guillaume Bécard[1], Marta Rodrigez-Franco[11], Thomas Ott[11,12], Justine Bertrand-Michel[8], Giles ED Oldroyd[3,13], Péter Szövényi[5,6], Marcel Bucher[2], Pierre-Marc Delaux[1]*

Correspondence to: pierre-marc.delaux@lrsv.ups-tlse.fr

This PDF file includes:

Materials and Methods
Figs. S1 to S18
Tables S2, S6, S7

Other Supplementary Materials for this manuscript include the following:

Table S1. GO term enrichment during AMS
Table S3. Differentially expressed genes in *Marchantia paleacea* during AMS
Table S4. Species sampling used for phylogenies
Table S5. Species sampling used for Orthofinder
Table S8. List of orthogroups

Materials and Methods

Biological material

Marchantia paleacea ssp. paleacea was kindly provided by David Beerling and Katie Field (Sheffield University, UK). *Marchantia paleacea ssp. diptera* was kindly provided by Masaki Shimamura (Hiroshima University, Japan). *Rhizophagus irregularis* DAOM197198 was obtained from Agronutrition, France.

Growth conditions

M. paleacea thalli were grown from cut thalli apex in 7*7*8 cm pots (5 plants per pot) on zeolite (0.5-2.5mm) at a 16h/8h photoperiod at 22°C/20°C and watered weekly with Long Ashton solution containing 15 µM of phosphate (*26*).

Marchantia paleacea ssp. paleacea genome sequencing and assembly

M. paleacea gametophytes were grown in axenic culture on BCD medium (9 cm diameter petri dishes) in a Panasonic growth chamber (60% Relative humidity, continuous light, 300 µEm2sec-1, 22 C) for about 4-8 weeks. We extracted DNA from the *M. paleacea* plant tissues using a modified CTAB method (*27*) by fishing out the precipitated DNA and washing it twice with ethanol. DNA was resuspended in TE at room temperature. DNA was QC-ed and quantified on a femtopulse machine (Femtopulse Sytems, Agilent) and four MinION libraries were prepared with the ligation kit (SQK-LSK109) each with approximately two micrograms of input DNA. Four MinION flow cells (FLO-MIN106) were ran for 72 hours, each with a separately prepared library, using a GridionX5 (Oxford Nanopore Technologies) at the Functional Genomics Center Zurich (FGCZ). Nucleotide base calling was carried out using the "high accuracy base calling" method in MinKNOW (v3.6.5) (https://community.nanoporetech.com) including Guppy (v3.2.10) (https://community.nanoporetech.com)

All reads shorter than 1kb were discarded and the minimap2 (v2.17) (*28*) - miniasm (v0.3-r179) (*29*) pipeline was used to assemble the raw reads into contigs with defaults parameters. We polished the resulting assembly with Racon (v1.4.3) (*30*) using only the ONT reads in two rounds. Finally, we carried out three rounds of pilon (v1.23) (*31*) polishing using the --fix indels and --haploid switches.

To remove scaffolds containing potential DNA from contaminants we blasted the scaffolds both against the National Center for Biotechnology Information (NCBI) nucleotide database (blastn) and the full uniprot (blastp) using an e-value threshold of 10-4. We also mapped the Illumina data to the genome (bowtie2 v2.3.5) (*32*) and ran BlobTools (v1.1) (*33*) to assign scaffolds to taxonomic groups based on Illumina sequencing depth and blast hits. After using blobtools taxify command, we removed all scaffolds that were classified as non-Streptophyte origin.

Annotation of protein coding and non-protein-coding gene models of the Marchantia paleacea ssp. paleacea genome

Gene models were predicted by the eukaryotic genome annotation pipeline *egn-ep* (http://eugene.toulouse.inra.fr/Downloads/egnep-Linux-x86_64.1.5.1.tar.gz) using trained statistical models adapted for plants (http://eugene.toulouse.inra.fr/Downloads/WAM_plant.20180615.tar.gz). The pipeline automatically manages probabilistic sequence model training, genome masking, transcript and protein alignments computation, alternative splice sites detection and integrative gene modelling by the EuGene software release 4.2a (*34*) http://eugene.toulouse.inra.fr/Downloads/eugene-4.2a.tar.gz).

Three protein databases were used to detect translated regions: i) the proteome of *Arabidopsis thaliana* TAIR10 version (*35*); ii) Swiss-Prot - October 2016, and iii) the proteome of *Medicago truncatula* A17 version 5 annotation release 1.6 (*36*) https://medicago.toulouse.inra.fr/MtrunA17r5.0-ANR/). Proteins similar to REPBASE (*37*) were removed from the three datasets (to avoid the integration of TE related proteins in the training steps). Chained alignments spanning less than 50% of the length of the database protein were removed. The proteome of *M. truncatula* was used as a training proteome by EuGene.

Two input transcripts for EuGene were used. One transcriptome based on mapping predicted from the RNAseq (Supplementary Material: Transcriptomic analyses) data of the time course (i.e. 3,5 and 8 weeks post-inoculation) in mock and mycorrhizal contexts and the mapping of three RNAseq samples data of *M. paleacea* from the NCBI [ERR2040988 and ERR2040989 from 1KP initiative, (*38*) and SRR5196885 from (*39*)]. Transcripts were predicted using Stringtie v2.1.4 (*40*) with *-f 0.8* on each sample. All 21 gtf sample files were merged together using *stringtie --merge* with standard options. Transcripts fasta file was generated using *gffread* v0.11.6 (*41*) with *-w* option. We also predicted a second transcriptome from the same RNAseq samples *de novo* using *DRAP*

v1.92 pipeline (http://www.sigenae.org/drap, (*42*). *runDrap* was used on the 21 samples applying the *Oases* RNAseq assembly software (*43*). *runMeta* was used to merge assemblies without redundancy based on predicted transcripts with fpkm 1. This transcriptome was employed as a training transcriptome by EuGene. Finally, 18768 protein-coding genes, 352 tRNAs, 77 rRNAs and 1507 ncRNAs were annotated in the annotation release 1.

Genome assembly, annotation file and gene models are available through a genome browser at https://bbric-pipelines.toulouse.inra.fr/myGenomeBrowser?browse=1&portalname=Marchantia_paleacea_PacBio&owner=cyril.libourel@lrsv.ups-tlse.fr&key=C3cEr7ro with LOGIN: pierre-marc.delaux@lrsv.ups-tlse.fr and PASSWORD: JdL0DpCk (Genome Browser access will be publicly shared upon publication).

Genome assembly statistics	
# contigs	192
Largest contig	8438675
Total length	250803610
GC (%)	40,14
N50	2390877
N75	1162399
L50	33
L75	71
# N's per 100 kbp	0
Genome annotation statistics	
Gene	
Count	18698
Average Length	3763,17
Median Length	2774
Average Coding Length	1081,17
Median Coding Length	753
Ave Exons Per Gene	4,23
Med Exons Per Gene	2
exon	
Count	79056
Average Length	255,71
Median Length	143
Intron	
Count	60361
Average Length	347,4
Median Length	235
UTR3	
Count	19247
Average Length	694,17
Median Length	535
UTR5	
Count	21337
Average Length	572,4
Median Length	384

BUSCO statistics	**eukaryota_odb09**	
	Absolute number	%
Complete BUSCOs (C)	261	86,2
Complete and single-copy BUSCOs (S)	249	82,2
Complete and duplicated BUSCOs (D)	12	4
Fragmented BUSCOs (F)	3	1
Missing BUSCOs (M)	39	12,8
Total BUSCO groups searched	303	

Table S2. *Marchantia paleacea* genome statistics.

Transcriptomic analyses

For the mycorrhizal time course, pots were inoculated one week after apex transfer with 1000 sterile spores of the AMF *Rhizophagus irregularis.* Inoculated and mock treated plants were harvested after three, five and eight weeks. For each condition, three samples were prepared from 3 independent experiments.

For the overexpressor experiment, three independent lines expressing *HA:MpaWRI* and *HA:MpaRAD1* or transformed with an empty vector control were harvested 5 weeks after transfer. Plants of each pot were pooled in a single sample, flash-frozen and stored at -70°C until extraction. TRI-reagent (Sigma) extraction was performed according to supplier's recommendation on ~100mg of ground frozen thalli. Around 2µg of RNA was treated with RQ1 DNase (Promega, USA) and sent for sequencing to Genewiz (Leipzig, Germany). Libraries were prepared with the NEBnext ultra II RNA non-directional kit for the mycorrhizal time course samples and NEBnext ultra II RNA directional kit for overexpressor lines samples.

The raw fastq paired-end reads were cleaned by removing the adapters and the low-quality sequences using cutadapt (v2.1, Martin 2011) and TrimGalore (v0.6.5, https://github.com/FelixKrueger/TrimGalore) with *-q 30 --length 20* options. The cleaned reads were mapped against the *Marchantia paleacea* genome using HISAT2 (v2.1.0, (*44*) with *--score-min L,-0.6,-0.6 --max-intronlen 10000 --dta* options and *--rna-strandness RF* option for RNAseq data from overexpressor. Duplicated reads were removed using SAMtools (v1.9, (*45*) *markdup* command. Gene counts were estimated with HTSeq-count (v0.11.2, (*46*) using the following options *-a 10 -r name -s no -m union --nonunique none* against the genome annotation of

Marchantia paleacea generated in this article. The Integrative Genomics Viewer software (IGV v2.5.0, (*47*) was used to visualize mapped reads.

Differentially expressed genes (DEGs) were identified using the R package edgeR (*48*, *49*) to compare the two *Marchantia paleacea* over-expressors against the empty vector with adjusted p-value (FDR method) < 0.05 (Table S3). The mycorrhizal samples (*i.e.* Myc) were compared to the mock (i.e. NM) treatment at the same time point (3wpi vs 3wpi, 5wpi vs 5wpi and Mock_8wpi vs 8wpi). The biological replicate 1 of Myc_3_wpi was removed from DEGs analysis because this sample was clustered with non-mycorrhizal samples.

Cross-referencing orthogroups and RNAseq

To compare and cross-reference the genes upregulated during the symbiosis across land plants, orthogroups were calculated with OrthoFinder v2.3.7 (*50*) for 11 species including six species with available expression data during arbuscular mycorrhiza symbiosis (Table S5, Table S8). Due to the disparity in the methods for identifying differentially expressed genes across the different publications, differentially expressed genes have been taken as published by the different studies without recalculation. For each orthogroups, the number of species having at least one gene up-regulated in the orthogroup was calculated. We also merged orthogroups of the five angiosperms species to the *M. paleacea* genes (Table S3).

GO term enrichment

To compare the gene ontology processes involved during the symbiosis across land plants, GO terms from InterProScan analysis of each of the six species were extracted using homemade R script. Based on genes significantly (FDR>0.05) up-regulated in mycorrhizal conditions, we estimated GO term enrichment for Biological Processes and Molecular Functions using *topGO* R package with fisher's statistic employing the *weight01* algorithm (Table S1, https://bioconductor.org/packages/release/bioc/html/topGO.html).

Functional annotation of protein-coding genes

Protein coding genes from the six species with available expression data during arbuscular mycorrhiza symbiosis were annotated using InterProScan v5.29-68 with the following options --*disable-precalc --goterms --pathways -iprlookup* (*51*).

Homologs identification & Phylogenetic analysis

Homologs of Fat, WRI, GPAT and KASI genes were retrieved from a database containing 89 species covering the main Viridiplantae lineages (Table S4). In addition to available genomic data, coding-sequences predictions from the 1KP data were added to empower the sampling in the Bryophytes and Charophytes lineages (*8*, *38*). Searches were performed using the respective *M. truncatula* (WRI, KASI, Fat) and *Arabidopsis thaliana* (GPAT) proteins with the tBLASTn v2.9.0+ algorithm (*52*), and an e-value threshold of 1e-10. Sequences were aligned using MAFFT v7.313 (*53*) and resulting alignments cleaned with trimAl v1.4.1 (*54*) to remove positions with more than 80% of gaps before being subjected to Maximum Likelihood (ML) analysis wit IQ-TREE v1.6.7 (*55*) . Prior to ML analysis, the best-fitting evolutionary model was tested using ModelFinder (*56*) and selected according to the Bayesian Information Criterion. Support for nodes was tested using 10,000 replicates of SH-aLRT and UltraFast Bootstraps (*57*, *58*). Spurious sequences corresponding to putative remnants, pseudogenes or miss-annotated genes were removed and a tree rebuilt following the method presented above. Trees were visualized and annotated through the iTOL v5.6.3 (*59*).

Phylogenies of STR, STR2 and RAD1, are available in (*8*)

	New ID	ID from Radakrishnan et al. 2020
MpaSTR	Marpal_utg000139g0174371	Marpal_scaff13491_FGenesh2.1
MpaSTR2	Marpal_utg000110g0155891	Marpal_scaff5479_FGenesh8.1
MpaRAD1	Marpal_utg000070g0119331	Marpal_scaff8805_FGenesh3.1
MpaRAD1-like	Marpal_utg000071g0120271	Marpal_scaff3964_FGenesh2.1

Domains and motifs identification

For each of the investigated genes, proteins and their homologs were subjected to HMMSEARCH from the HMMER v3.3 package (*60*) to find functional domains. Searches were performed using an e-value threshold of 1e-04 for both full and domain hits. Functional domains were mapped on the respective phylogeny. In addition, previously identified motifs (*61*) allowing the distinction of

the different AP2 subfamilies were searched in the homolog proteins of WRI using the motifSearch package (available at: https://github.com/jeankeller/motifSearch, Fig. S6, Table S6).

motif_name	motif_regex
M1	DK[NSG][STC]WN
M1b	DKX[STC]WN
M2	L[SGE]LLL[QR]S[STP][MKY]F
M2b	L[SGE]LLL[QR]SXXF
M3	F[PE][DE][DN][IV]XF[DGE]
M4	[ATM][VSLA]L[RH]NL[MI]GL[DES]
M5	Q[ML]LXQQQ
M6	[QH][SPA]L[STA]LM[SNG]X[GTA]
M7	DV[KED]XI[LCM][SEA]S[ST][SN][TLS]
M8	DGSLCIMEA[LF]
M9	G[GDE]XA[KR]RXK
M10	E[SPT][RK]KR[GA]X[GVN][KR]
M11	[HQ]NF[FL]QX
M12	INVN[LY][PSM][YP]
M13	[KR][EQK][PSA][VS][PH]R
euANT1	DNSC[RK][KR]
euANT2	NNWL[GA]FSLS[PSN]
euANT3	PK[LV]EDFLG
euANT4	KSX[DE][TS]FGQRT

Table S6. AP2 motifs from Dipp-Álvarez *et. al* 2019 (61)

GRAS transcription factors screening

To investigate *RAD1* putative orthologs in Charophytes, GRAS transcription factor family were screened from all Charophytes species with available genome, all Bryophytes used in this publication and the newly sequenced moss *Ceratodon purpureus* (*62*) as well as species representative of Angiosperms *(A. commosus, A. thaliana, A. trichopoda, M. truncatula, O. sativa),* In addition, 3 species of Chlorophytes algae (*V. Carteri, D. salina, C. reinhardtii*) were added to confirm the absence of GRAS transcription factors in this lineage. Proteomes of each species were screened using HMMSEARCH module from the HMMER 3.3 package (*60*) and the HMM model of GRAS domain PF03514.15 from PFAM database (*63*). Searches were performed with an global and domain e-value threshold of 1e-04. In parallel, BLASTp v2.9.0+ (*52*) search of

all *A. thaliana* annotated GRAS transcription factors was performed on all proteomes with an e-value threshold of 1e-05. Proteins identified through HMMSEARCH and BLASTp searches were merged prior alignment using Clustal Omega v1.2.4 (*64*) with default parameters. Resulting alignment was trimmed using trimAl v1.4.1 (*54*) to clear positions with more than 60% of gaps. Cleaned alignment was subjected to ML analysis following the procedure described above.

Cloning

Constructs were generated in the modular GoldenGate system (*65*) starting from domesticated (for the restrictions sites DraIII, BpiI, BsaI and BsmBI) and synthetized level 0 parts. Digestion-ligation reactions (37°C - 2 min, 16°C - 5 min) for 25 cycle, 5 min - 37°C, 5 min - 16°C) were made with T4 DNA ligase (New England Biolabs, UK) and the restriction enzymes BsaI-HF (New England Biolabs, UK) or BpiI (Thermofisher, USA) for level 1 or level 2 respectively.

To overexpress *HA:MpaWRI* and *HA:MpaRAD1*, we used the *M. polymorpha* constitutive promoter EF1a (*66*). Expression of *HA:MpaWRI*^{SRDX} under the pMpoEF1a led to a drastic reduction of the number of plant transformants, we therefore used an inducible bipartite dexamethasone inducible promoter (*pMpoEF1a:GVG, pGal4:HA: MpaWRI*^{SRDX}). All transgenic plants containing this construct showed the phenotype presented in Fig. S11 even in absence of induction, indicative of a promoter allowing constitutive expression.

For transient expression in *N. benthamiana*, *MpaWRI* and *MtWRI5b* were expressed under the *A. thaliana Ubiquitin 10* promoter.

All GoldenGate level 0 used in this study are held for distribution in the ENSA project core collection (https://www.ensa.ac.uk/).

To produce GST-MpaWRI, *MpaWRI* was cloned in pDONR201 and subsequently introduced in pGEX-6P-1 using BP clonase and LR clonase (Invitrogen, USA) following supplier's protocol.

Vector list and maps are available in Table S7.

L2 plasmids - Overexpressors	
pL2B-HYG-EmptyVector	
pL2B-HYG-pMpoEF1a-HA:MpaRAD1-T35S	MpaRAD1: Marpal_utg000070g0119331
pL2B-HYG-pMpoEF1a-HA:MpaWRI-T35S	MpaWRI: Marpal_utg000074g0124251
pL2B-HYG-pGal4-HA:MpaWRI:SRDX-pEF1a-GVG	

L2 plasmids - GUS reporter lines	
pL2B-HYG-pMpaRAM2A_2kb-t35S	MpaRAM2A: Marpal_utg000010g0023111
pL2B-HYG-pMpaRAM2B_2kb-t35S	MpaRAM2B: Marpal_utg000148g0178951
pL2B-HYG-pMpaRAM2C_2kb-t35S	MpaRAM2C: Marpal_utg000108g0154471
pL2B-HYG-pMpaRAM2C_1.5kb-t35S	
pL2B-HYG-pMpaRAM2C_1kb-t35S	
pL2B-HYG-pMpaRAM2C_798bp-t35S	
pL2B-HYG-pMpaRAM2C_500bp-t35S	
pL2B-HYG-pMpaRAM2C_250bp-t35S	
pL2B-HYG-pMpaWRI_3kb-GUS-t35S	
pL2B-HYG-pMpaRAD1_2kb-GUS-T35S	
pL2B-HYG-pMtRAM2_2.3kb-GUS-t35S	

L2 plasmids - Lipid tranfer	
pL2B-HYG-pMtRAM2-UcFatB-t35S	UcFatB: from Luginbuehl et al. 2017

L2 plasmids - CRISPR
pL2V-HYG-pMpoEF1a:AtCas9-pMpaU6-1:MpaWRI-sg3-sg4

EMSA
pGEX-6P-1_MpaWRI

Table S7. List of constructs

CRISPR/Cas9

To generate mutants of *MpaWRI*, plants were transformed with constructs containing a *Arabidopsis thaliana* codon optimized *Cas9* (*67*) under the *MpoEF1a* promoter and two guide RNA under the control of *M. paleacea* U6 promoter. Seven edited lines resulting from one guide pair (sg3 AGTCGCTCCCGCCGAAGCCA, sg4 TGCCACCGCTTTCTGCTCCT) were selected

for phenotyping (Mpa*wri-1* to Mpa*wri-7*). Two of those lines showed editions that modified the gene without impacting the open reading frame Mpa*wri-6*, Mpa*wri-7*, Fig. S13)

To generate mutants of WRI in *Marchantia paleacea ssp. diptera* (wri-8), plants were transformed with the construct containing Arabidopsis-codon-optimized Cas9 fused with *MpoEF1a* promoter and a guide RNA (GCGATCGCGCCGCGACAG) fused with *MpoU6-1* promoter. One edited line was selected for phenotyping.

Transformation of Marchantia paleacea

Gemmae of axenic *M. paleacea* were grown 4 weeks on ½ Gamborg media (Duchefa, NL) 1% bactoagar (Euromedex, FR). Around 30 plantlets per construct were blended for 15 seconds in 10ml OM51C (*68*). Blended material was transferred in a 250ml Erlenmeyer containing 40ml OM51C and grown at 200 RPM 20°C under continuous light. After 3 days, 200µl of a saturated *Agrobacterium tumefasciens GV3101* liquid culture, glutamine (5 g/L final) and acetosyringone (100µM final) were added and plants and bacteria were co-cultivated for 3 days. Thalli fragments were washed with water by decantation in 50ml tubes three times and selection was performed on ½ gamborg, 10 mg/L hygromycin, 200 mg/L amoxicillin (Augmentin). For *M. paleacea ssp. diptera* 500nM Chlorsulfuron (Wako) was added.

Mycorrhizal phenotyping

Thalli were harvested 5 weeks post inoculation and embedded in 6% agarose and 100 µm longitudinal sections were made using a Leica vt1000s vibratome. Sections were incubated overnight in 10% KOH to remove the mycorrhiza specific purple pigment, washed and stained 30min in 5% Sheaffer ink (USA), 5% acetic acid. For the images in figure 3, fresh sections were incubated overnight in PBS containing 1µg/ml WGA-FITC (Sigma) and pictures were taken with a Zeiss Axiozoom V16 microscope.

Phenotyping of wri-1 to 7 and two independent empty vector lines were performed in 3 independent experiment with anonymized lines ID. Phenotyping of wri-8, WRI^{SRDX} and corresponding wild types was performed twice.

TEM pictures

Vibratome sections were fixed in MTSB (*69*) buffer containing 2.5 % glutaraldehyde and 4% p-Formaldehyde under vacuum for 15 min and stored at 4°C in fixative solution until further steps. 1 mm^2 pieces of the fixed sections were dissected from the areas of tissue containing purple pigment, washed 5 times for 10 min each with buffer, and post-fixed with 1% OsO_4 in H_2O at 4°C for 2 h. After 5 times washing (10 min each) with H_2O at room temperature, the tissue was *in bloc* stained with 1% Uranyl Acetate for 1 h in darkness. Samples were washed 4 times (10 min each) in H_2O and dehydrated in EtOH/H_2O graded series (30%, 50%, 70%, 80%, 90%, 95% 15min each). Final dehydration was achieved by incubating the samples twice in absolute EtOH (30 min each) followed by incubation in dehydrated acetone twice (30 min each). Embedding of the samples was performed by gradually infiltrating them with Epoxy resin (Agar 100) mixed with acetone at 1:3, 1:1 and 3:1 ratios for 12 h each, and finally in pure Epoxy resin for 48 h with resin changes every 12 h. Polymerization was carried out at 60°C for 48 h. Ultrathin sections of approximately 70 nm were obtained with a Reichert-Jung ultra-microtome and collected in TEM slot grids. Images were acquired with a Philips CM 10 transmission electron microscope coupled to a Gatan BioScan 792 CCD camera at 80 kV acceleration voltage.

GUS staining

Plants were harvested 5 weeks post inoculation, washed, submerged in GUS buffer (10mM phosphate buffer pH 7, 0.5mM EDTA, 0.5mM potassium ferricyanide, 0.5mM potassium ferrocyanide and 0.25mg/mL X-Gluc (Euromedex, France)), vacuum infiltrated for 10min and incubated overnight at 37°C. Thalli were embedded in 6% agarose and 100 µm transversal sections were made using a Leica vt1000s vibratome. Sections were incubated overnight in 10% KOH before observation.

Observation of promoter:GUS pattern of expression was performed at least 3 times.

Quantitative RT-PCR

RNA of HA:MpaWRI, HA:MpaWRI^{SRDX} or empty vector control plants was extracted as seen in "Transcriptomic analysis". Reverse transcription was performed using M-MLV (Promega, USA) on around 1µg of RNA and qPCR was performed on 5x diluted cDNA in a LC480 thermocycler (Roche) with SYBR Green (Sigma). Relative expression values were calculated using reference genes MpaEF1a and MpaAPT2. Two independent experiments were performed with similar results.

MpaAPT2 qPCR F	GGGTACACTTGCTGCAGGAA
MpaAPT2 qPCR R	CTCACGGCCCTTTAGATCCG
MpaEF1a qPCR F	AATGTGTTGAGCAGCTTGGC
MpaEF1a qPCR R	ACGTTCCAAGTACTCTCGAGC
MpaCBX2 qPCR F	CGTCGTACCTCTTCTCGAACC
MpaCBX2 qPCR R	AAGGATGCATAAGCCGACAG
MpaRAM2A qPCR F	GGAACCCTTGGCTTCGAATG
MpaRAM2A qPCR R	AATGACTTTTTCGGTGCTACACT
MpaRAM2B qPCR F	CGCTGAATTTCGAGTGCACG
MpaRAM2B qPCR R	CGTTTTCCGGGGGGCAA
MpaRAM2C qPCR F	ACCCTCAGCTTCGAGTGTAC
MpaRAM2C qPCR R	GCTTGCGCTCTTCTTCGG
MpaKasI-F qPCR	AACTCGTTCGGGTTTGGAG
MpaKasI-R qPCR	TCCTGCTAGCTCATCCCTTG

Transient expression of WRI in *Nicotiana benthamiana* leaves

Three weeks old *N. benthamiana* leaves were infiltrated with a mix of *A. tumefaciens* GV3101 containing *MpaWRI*, *MtWRI5b* or empty vector control and P19. Three one-cm leaf discs from three independent experiments were harvested for each construct 5 days post infiltration and freeze dried. Figure S10 show all data collected.

Marchantia triacylglycerol extraction

To analyze *M. paleacea* lipid accumulation during AMS, thalli were harvested 4 weeks post inoculation and dissected to retrieve the central and colonized part of the thallus. Samples were weighed and frozen in liquid nitrogen and stored at -80°C before analysis. This experiment was performed three times with similar results.

Neutral lipid profiling

Lipids corresponding from 75mg to 200mg of plant material were extracted according to Bligh and Dyer (*70*) in dichloromethane/methanol/water (2.5 :2.5 :2.1, v/v/v), in the presence of the internal standards : 4 μg of stigmasterol, 4 μg of cholesteryl heptadecanoate, 8 μg of glyceryl tri nonadecanoate. Dichloromethane phases were evaporated to dryness and dissolved in 30μl of ethyl acetate. 1μl of the lipid extract was analyzed by gas-liquid chromatography on a FOCUS Thermo Electron system using an Zebron-1 Phenomenex fused silica capillary columns (5m X 0,32mm i.d, 0.50 m film thickness) (*71*). Oven temperature was programmed from 200°C to 350°C at a rate of 5°C per min and the carrier gas was hydrogen (0.5 bar). The injector and the detector were at 315°C and 345°C respectively.

Fatty Acid Methyl Ester (FAME) analysis

Dissected thalli were extracted according to Bligh and Dyer (*70*) in dichloromethane/methanol/water (2.5 :2.5 :2.1, v/v/v), in the presence of the internal standards glyceryl triheptadecanoate (2μg). The lipid extracts were hydrolysed in KOH (0.5M in methanol) at 50°C for 30', and transmethylated in boron trifluoride methanol solution 14% (SIGMA, 1ml) and heptane (1ml) at 80°C for 1h. After addition of water (1ml) to the crude, FAMEs were extracted with heptane (3ml), evaporated to dryness and dissolved in ethyl acetate (20μl). FAME (1μl) were analyzed by gas-liquid chromatography (*72*) on a Clarus 600 Perkin Elmer system using a Famewax RESTEK fused silica capillary columns (30 m x 0.32 mm i.d, 0.25 m film thickness). Oven temperature was programmed from 110°C to 220°C at a rate of 2°C per min and the carrier gas was hydrogen (0.5 bar). The injector and the detector were at 225°C and 245°C respectively.

Lipid-transfer analyses

2-month-old gemmae containing MtRAM2$_{pro}$:UcFatB or wild type *M. paleacea* were transplanted in two-compartment pots containing equal parts of zeolite (1-2.5mm) and sand, separated by a 25 µm nylon mesh. The upper part of the pot was inoculated with 1000 sterile spores of *R. irregularis*. Thalli were harvested after 4 months and dissected to retrieve the central and colonized part of the thallus. Newly formed spores and hyphae in the lower part of the pot, below the nylon mesh, were recovered using a modified version of Fang et al. 1983 (*73*) isolation method. Samples were weighed and frozen in liquid nitrogen and stored at -80°C before analysis.

Figure 2B and C show results of 2 independent experiments.

Triacylglycerol quantification by Mass Spectrometry

Triglycerides profiling was done on dissected thalli at the Metatoul-lipidomic platform (Inserm UMR1048, Toulouse, France) certified to ISO 9001:2015 standards. The extracts obtained for neutral lipid analysis were kept at -20°C until the analysis by LC-HRMS. They were run on a Ultimate 3000 (ThermoFisher Scientific (Life Technologies SAS, Saint-Aubin, France) equipped with a Acquity BEH C8 column (2.1 x 100 mm, 1.7 µm, Waters SAS, Saint-Quentin-en-Yvelines) kept at 60°C. Mobile phases were composed of (A) Water LCMS Grade/Methanol LCMS Grade (60/40 ; v/v) with 10 mM ammonium formate and 0.1% formic acid in, and (B) Isopropanol LCMS Grade/Methanol LCMS Grade (90/10 ; v/v) with 10 mM ammonium formate and 0.1% formic acid. The flow rate is 0.4 mL/min. The gradient was as follow: 32% B at 0 min for 2.5 min; 45% B at 3.5 min; 52% B at 5 min; 58% B at 7 min; 66%B at 10 min; 70% B at 13 min; 80%B at 16 min; 85% B at 23 min; 90% at 26 min and 95% B at 28 min. The injection volume was 5 µL. High-resolution mass spectrometry detection was done on an Exactive mass spectrometer (ThermoFisher Scientific, Life Technologies SAS, Saint-Aubin, France) in an ESI positive mode and an acquisition in full scan mode from 250 to 1800 m/z. Parameter of the source were as follow : Sheath Gaz : 35psi, Auxiliary Gas Flow : 15Psi, Sweep as Flow Rate : 0 Psi, Spray voltage : 3500V, Capillary Temperature : 250°C, Capillary Voltage : 25V, Tube Lens Voltage : 120V, Heater Temperature : 350°C, Skimmer Voltage : 22V, Maximum Injection Time : 250ms. Relative quantification was carried out using the ion chromatogram obtained for each compound using 5 ppm windows with Trace Finder Quantitative Software (ThermoFisher Scientific, Life Technologies SAS, Saint-Aubin, France).

Lipid detection by microscopy

Approximately 0.5 mm hand-cut sections were made on fresh, mycorrhized, *M. paleacea* with a razor blade. The sections were stained with 1µg/mL Nile red dissolved in water for 30 minutes. Plant and fungal cell wall were counterstained with 1mg/mL calcofluor for one minute, then briefly rinsed in water before mounting, in water, between slice and coverslip for observation.

Images were acquired using a Leica SP8 confocal laser scanning microscope, driven with LAS X software with objectives 10x/0.3 dry, 25x/0.95 or 40x/0.8 water immersion objectives. The Nile red dye was excited with an Argon laser at 488nm and the fluorescence emission was detected with two photomultipliers between 550 and 585 nm and between 585 and 640 nm. Calcofluor was excited using UV laser at 405 nm and fluorescence detected using an emission window of 415 to 485 nm. This experiment was performed twice with similar result.

Purification of GST-MpaWRI Protein Expressed in *Escherichia coli*

E. coli BL21 (DE3) cells carrying pGEX-MpaWRI were grown at 28°C to OD_{600} of 0.4 to 0.6. Expression of the GST-MpaWRI protein was induced by 1 mM isopropyl-β-d-thiogalactopyranoside for 4 h at 37°C. Cells were harvested by centrifugation (8000 rpm for 15 min at 4°C) and then resuspended and sonicated in cold lysis buffer [50 mM NaH_2PO_4, 300 mM NaCl, 1 mM phenylmethylsulfonylfluoride, bacterial protease inhibitor cocktail (500 µl of cocktail solution per 20 ml of cell lysate containing 2g of *E. coli* cells, Sigma P8465), pH 8.0]. The cell lysate was centrifuged (16000 rpm for 30 min at 4°C), and the cleared extract was mixed with 2ml of pre-equilibrated glutathione-sepharose 4B (GE Healthcare GE17-0756-01). Then it was shaken gently on a rotary shaker for 60 min at 4°C. The lysate/glutathione-sepharose 4B mixture was loaded onto a column with capped bottom outlet. The column washing performed with 100 ml of washing buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM phenylmethylsulfonylfluoride) and then followed by washing with 10 ml of 50mM Tris pH 7.5. The GST-MpaWRI protein was eluted from the column using elution buffer (50mM Tris pH 7.5 and 10 mM glutathione). The eluted GST-MpaWRI protein fraction was further concentrated using amicon ultra-4 centrifugal filter unit (Millipore UFC801024 MM). Finally, protein concentrations were adjusted based on Bradford assays (Bio-Rad) using a BSA standard curve and then confirmed by SDS-PAGE and Coomassie Blue analysis.

Electrophoretic Mobility Shift Assays (EMSAs)

EMSAs were performed using 60-240 ng of purified GST-MpaWRI fusion protein incubated in binding buffer [30 mM Tris pH 7.5, 50 mM KCl, 5 mM MgCl2, 10% glycerol, 0.02 µg/µl BSA, 1 mM DTT, 1 mM EDTA, 0.05 µg/µl Poly(dI-dC), 75 nM dsDNA probe], on ice in darkness for 30 min and subsequently loaded onto 6% TBE-polyacrylamide gels. Samples were run for 45 min at 120 V at 4°C. However, the running time for samples containing long probes (C, B1, and B2) was 110 min. The gels were directly visualized using ChemiDoc™ MP imaging system (Biorad) using a cy5 filter. The following pair primers were used to generate C, B1, and B2 fragments as dsDNA probes using polymerase chain reaction: probe C ([Cy5]AGGCCCGCGCAAGCCCAGGCGCTA and CCTCCCCTGGGGCGGGGCGGGGCG), probe B1 ([Cy5]CGCCCCGCCCCGCCCCAGGGGAGG and GATGGAAATTACCATGACCTAGCAC) and probe B2 ([Cy5]GTGCTAGGTCATGGTAATTTCCATC and GACCGTTAGGCTCTTGTCATCAACAGACG). The AW1 ([Cy5]AGGCCCGCGCAAGCCCAGGCGCTA and TAGCGCCTGGGCTTGCGCGGGCCT), AW2 ([Cy5]CGCCCCGCCCCGCCCCAGGGGAGG and CCTCCCCTGGGGCGGGGCGGGGCG), CTTC ([Cy5]CGTCTGTTGATGACAAGAGCCTAACGGTC and GACCGTTAGGCTCTTGTCATCAACAGACG), B1a ([Cy5]CGACCATCCCGGAGCAGCGATTCCT and AGGAATCGCTGCTCCGGGATGGTCG), and a mutated probe mB1a ([Cy5]CGACCATGGCGGAGCAGCGAAAACT and AGTTTTCGCTGCTCCGCCATGGTCG). For competition experiments, an unlabeled B1a probe was prepared (CGACCATCCCGGAGCAGCGATTCCT and AGGAATCGCTGCTCCGGGATGGTCG). Probes were prepared by directly mixing the corresponding primers.

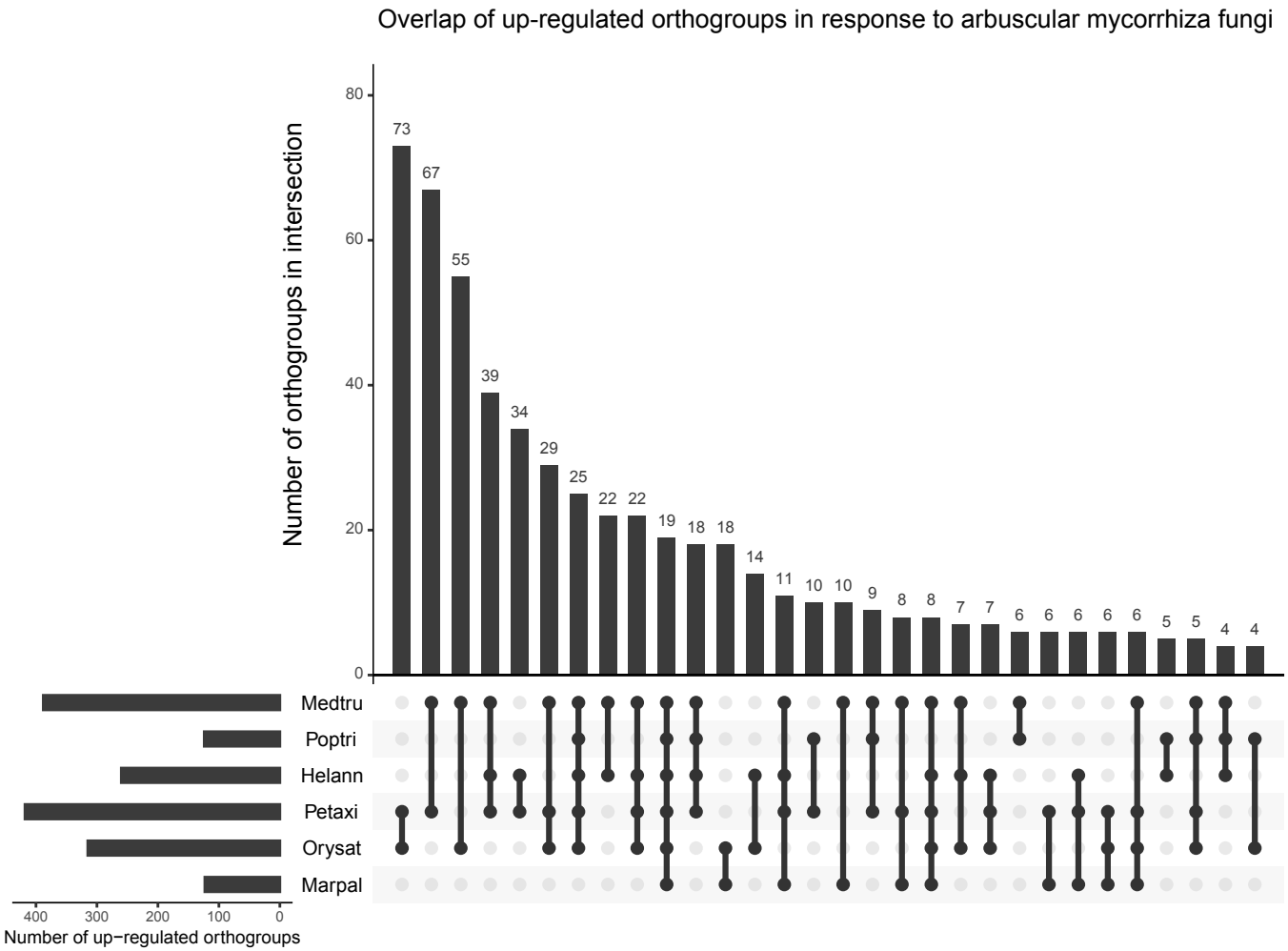

Fig. S1. Comparison of transcriptomic response across species. Matrix of intersections between orthogroups containing upregulated genes. 19 orthogroups are upregulated in all 6 land plant species and 25 are shared among the 5 angiosperms only.

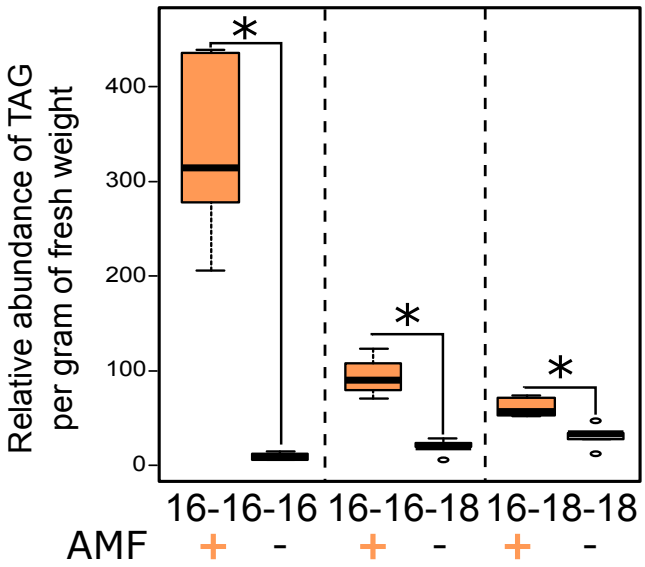

Fig. S2. Quantification of triacylglycerol in *Marchantia paleacea*. Relative amount of triacylglycerol with C18 and/or either one, two or three palmitic acid (C16) from dissected thalli of *M. paleacea*. AMF +: thalli inoculated with the AMF *Rhizophagus irregularis*; AMF -: mock-treated plants.* indicates p-value <0.05 according to a Mann-Whitney statistical test.

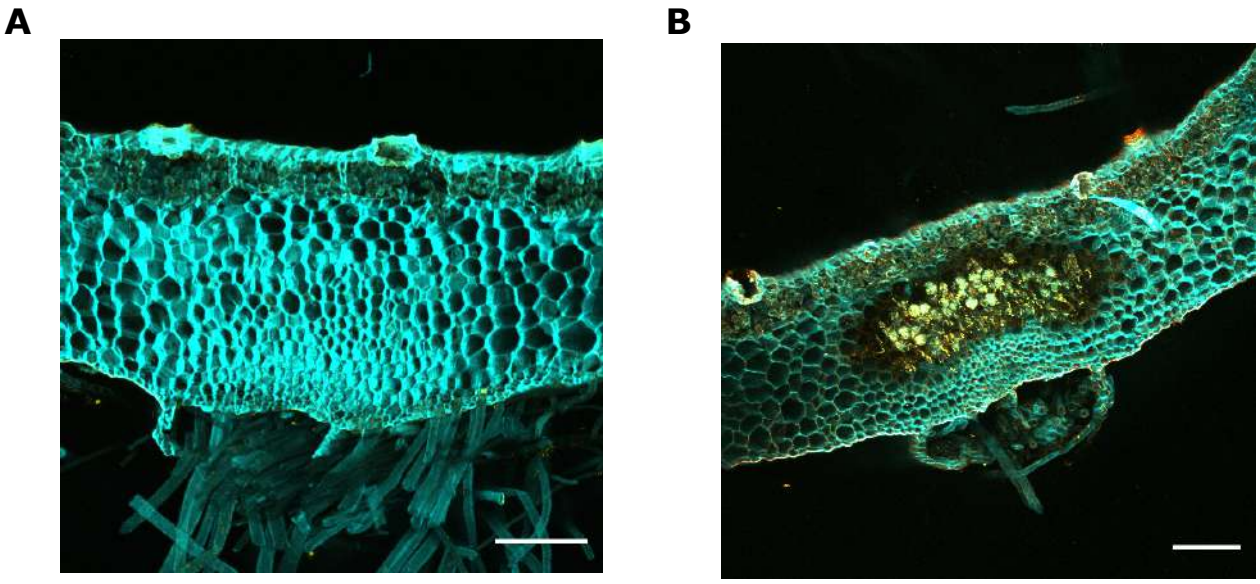

Fig. S3. Nile-red staining of *Marchantia paleacea*. Confocal imaging of mock-inoculated (**A**) or AMF-inoculated (**B**) *M. paleacea* thallus. Lipids were stained with Nile red (yellow) and cell wall with calcofluor (Blue). Bar = 200 µm.

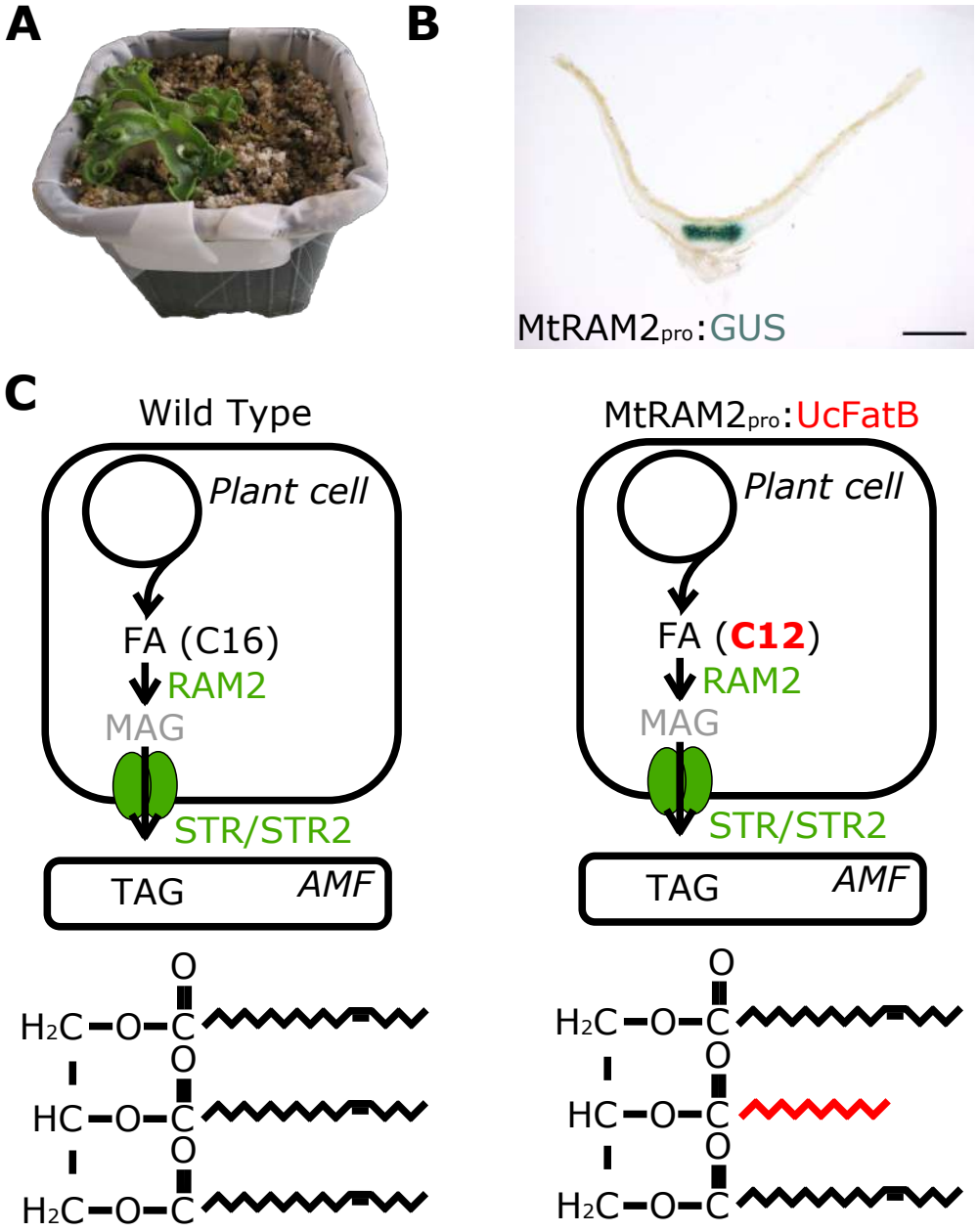

Fig. S4. Experimental setup for Fig. 2B. (**A**) *Marchantia paleacea* wild type or *MtRAM2$_{pro}$:UcFatB* were grown in double compartment pot system separated by a 25µm mesh, the lower part is plant free. (**B**) The *Medicago truncatula* RAM2 promoter used to drive the UcFatB expression shows arbuscule specific expression in *M. paleacea*. Bar = 1 mm (**C**) Rewiring of the putative lipid transfer pathway. Plant fatty acids are converted into monoacylglycerol (MAG) by RAM2, translocated to the peri-arbuscular space by half ABC transporters STR/STR2, taken up by the fungus and stored as triacylglycerol (TAG). Ectopic expression of *Umbellularia californica FatB* in plant arbuscule containing cells produces unique lauric fatty acids (C12) that can be measured in distant fungal hyphae and spores.

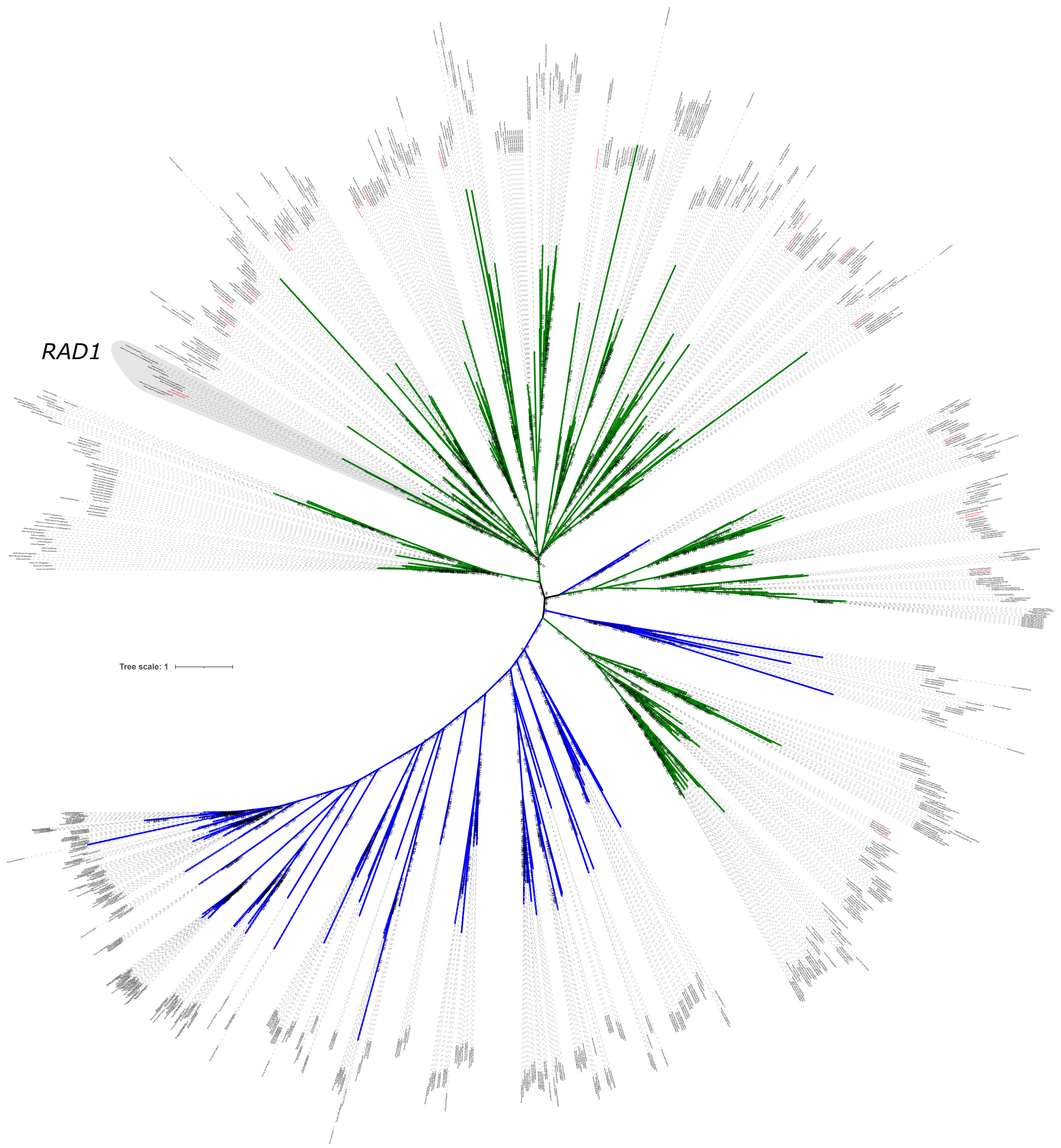

Fig. S5. Maximum likelihood tree of GRAS transcription factor family (model: JTT+F+R9; log-likelihood: -247748.8869). Branches colored in blue and green indicate algae (Chlorophytes and Charophytes) and land plants lineages respectively. *RAD1* clade is highlighted in grey.

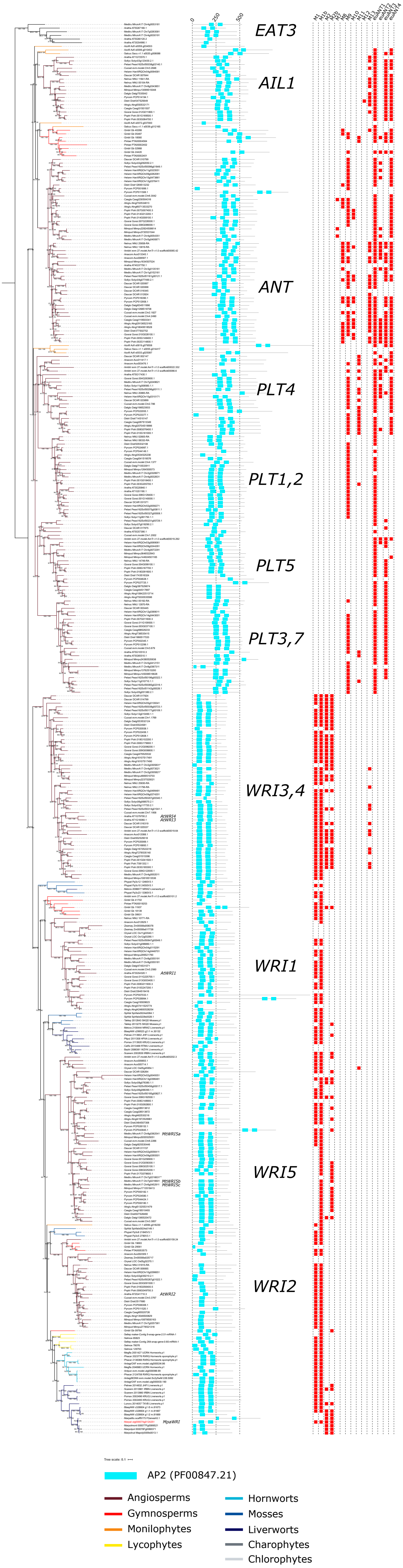

Fig. S6. Maximum likelihood tree of *WRI* and *PLT* clades (model: GTR+F+R8; log-likelihood: -275564.5304) Tree is rooted using representative sequences from EAT3 subfamilies of AP2. AP2 functional domains (PF00847.21) are indicated by light blue boxes at the right of the tree. Their size and position is proportional to the domain and protein length respectively. Motifs found in protein sequences are mapped in the columns at the extreme right of the tree. Red square indicates presence of the motifs while empty square denotes the absence of the motifs. List of motifs and their associated regular expression used for the search is available in table S5. Branches are colored according to the main plant lineage as described at the bottom of the tree. Pseudogene of *M. truncatula* WRI5 gene is indicated with ** and two gene models that normally form a single gene are marked with *.

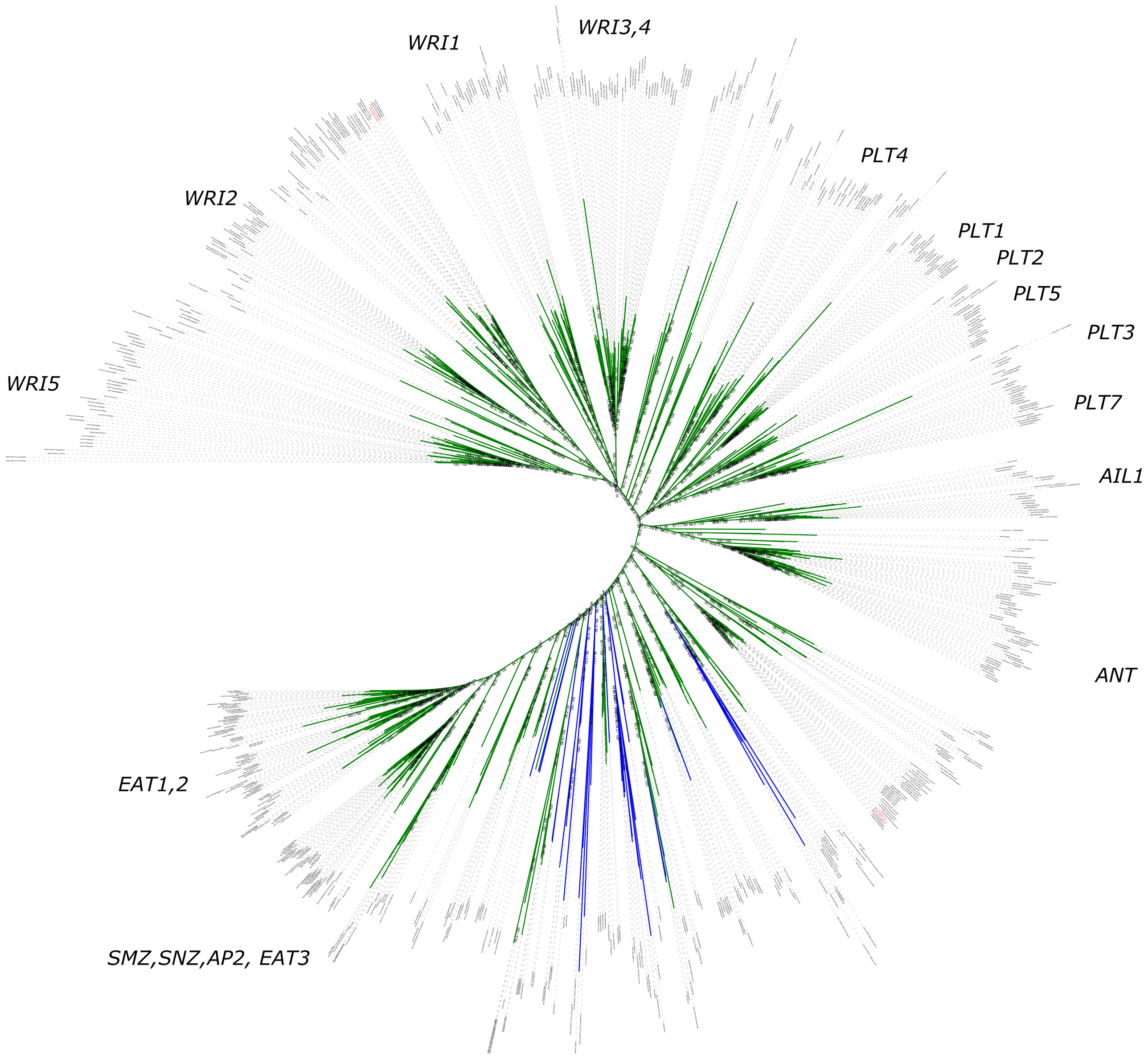

Fig. S7. Maximum likelihood tree of *AP2* clades (model: GTR+F+ASC+R9; log-likelihood: -518081.4578). Branches colored in blue and green indicate algae (Chlorophytes and Charophytes) and land plants lineages respectively. Clade names are indicated around the tree according the *A. thaliana* nomenclature.

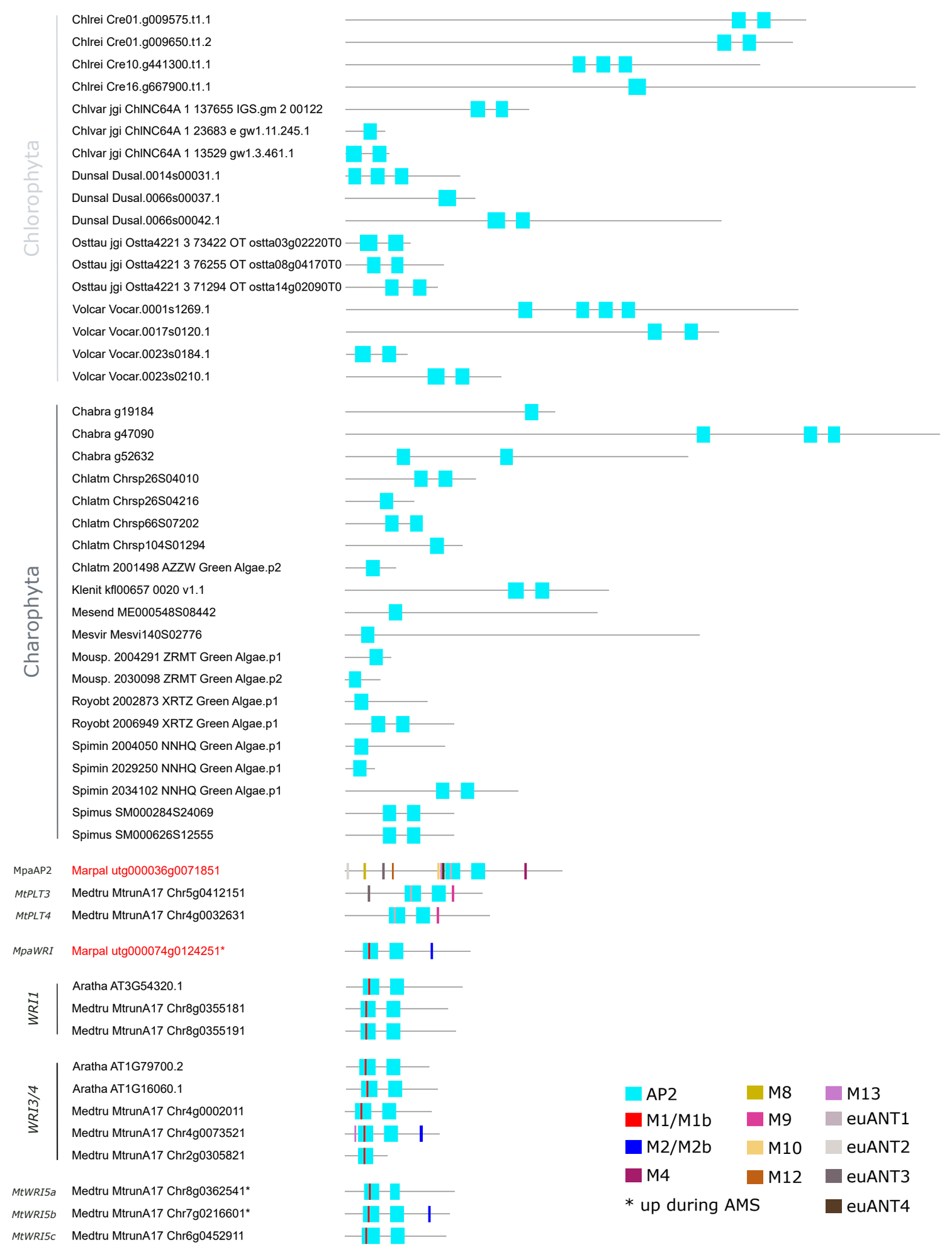

Fig. S8. Protein motifs in WRI. Characteristic motifs of WRI proteins were identified and mapped onto proteins of representative species from WRI clade, its closest paralog PLT and all Chlorophyta and Charophyta. Proteins backbone is represented by horizontal black lines, with their length being proportional to the protein length.
AP2 functional domains (PF00847.21) are indicated by light blue boxes, their length and position are proportional to the motif and protein size respectively. Horizontal colored lines represent motifs previously published. Color is described in the figure legend. The position of motifs is relative to the protein size. Asterisks indicate up-regulated genes during AMS.

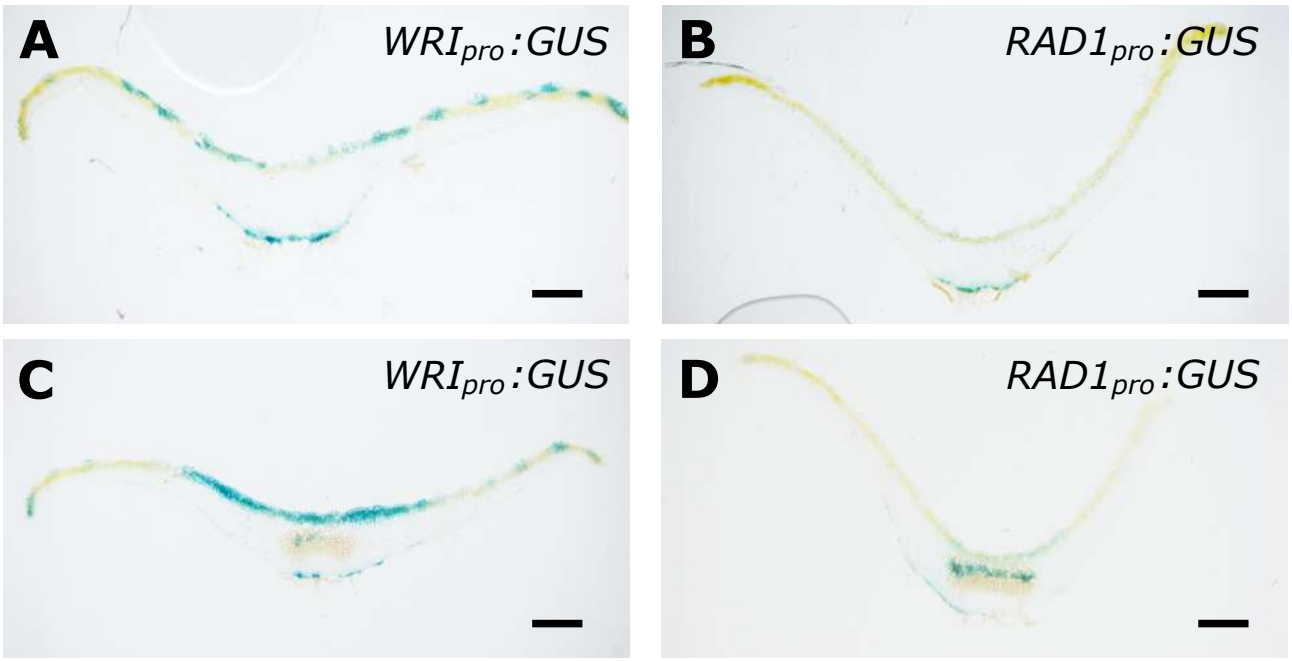

Fig. S9. *MpaWRI* and *MpaRAD1* are expressed in the central arbuscule containing zone. Transversal sections of thalli expressing *MpaWRI*$_{pro}$*:GUS* (**A, C**) and *MpaRAD1*$_{pro}$*:GUS* (**B, D**) In mock (**A, B**) or mycorrhizal condition (**C, D**). Four independent lines for each construct show a similar pattern of expression. Bar = 0.2 mm

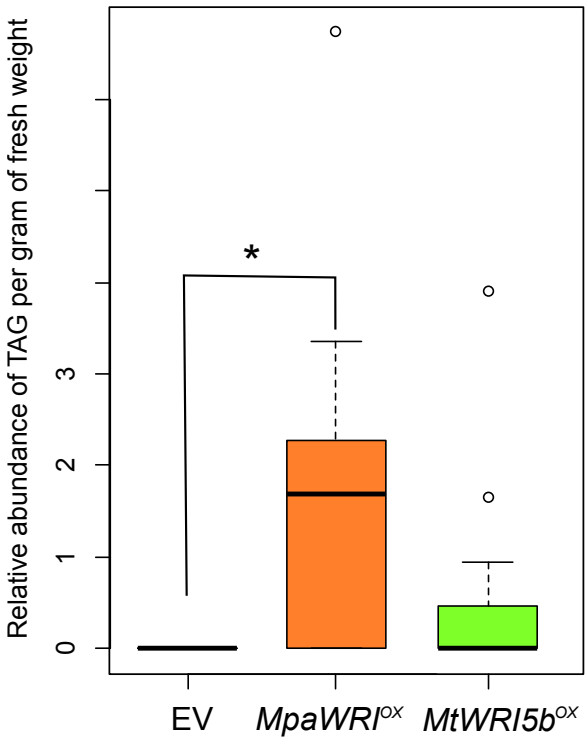

Fig. S10. Quantification of triacylglycerol in *Nicotiana benthamiana.* Relative quantification of total tri-acylglycerols in *Agrobacterium tumefascens* infiltrated *N. benthamiana* leaves expressing *Marchantia paleacea WRI* (*MpaWRI*), *Medicago truncatula WRI5b* (*MtWRI5b*) or transformed with an empty vector (EV).
* indicates p-value <0.05 accordind to a Mann-Whitney statistical test.

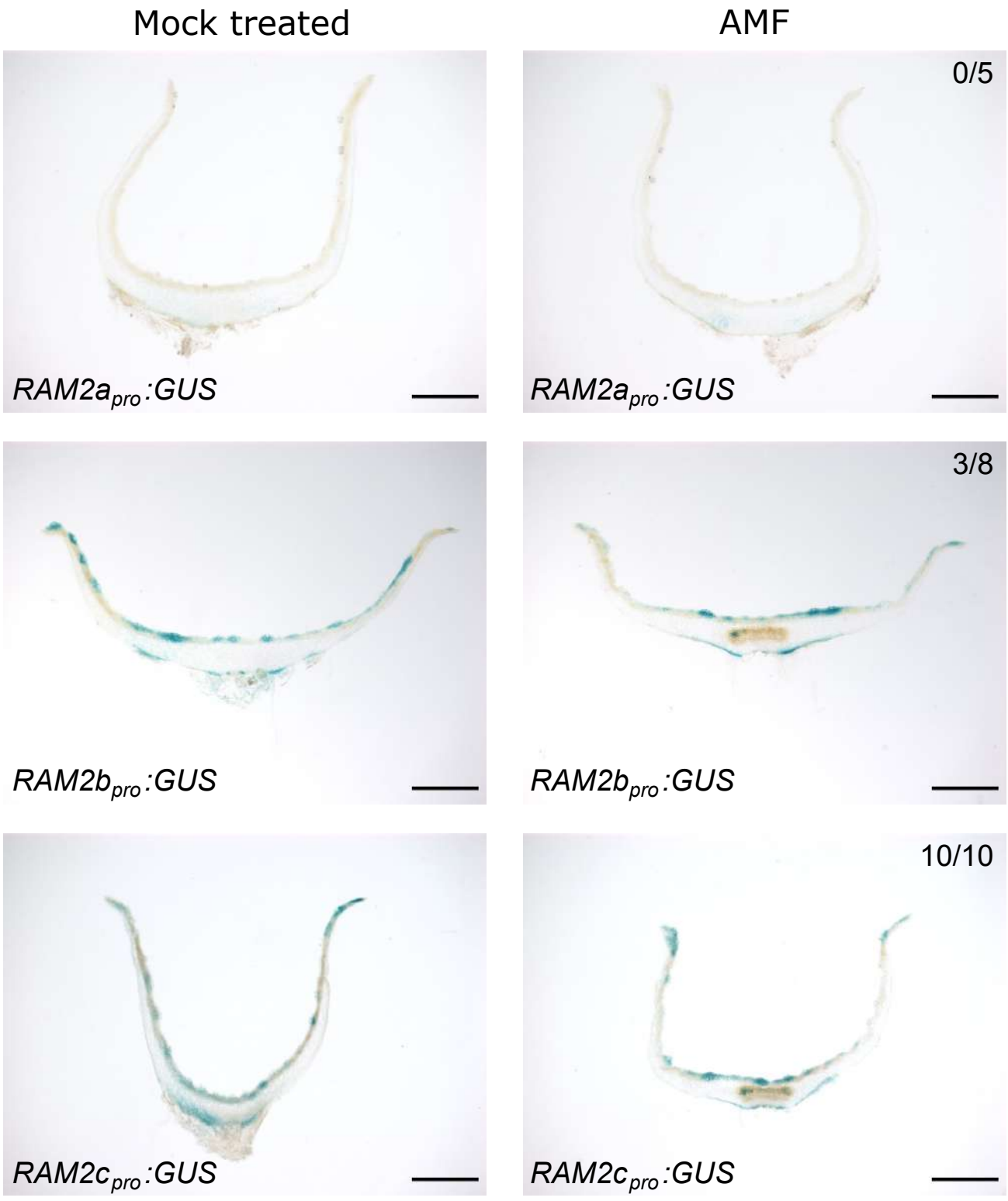

Fig. S11. Expression pattern of the *Marchantia paleacea RAM2 paralogs.* Expression pattern using promoter GUS fusions of *MpaRAM2a* (Marpal_utg000100g0149511, *MpaRAM2a:GUS*), *MpaRAM2b* (Marpal_utg000148g0178951, *MpaRAM2b:GUS*) and *MpaRAM2c* (Marpal_utg000108g0154471, *MpaRAM2c:GUS*). The percentage of colonized thalli showing GUS staining in the arbuscule-containing area is indicated in the top-right corner of the photos. Mock treatment: non-colonized thalli; AMF: thalli colonized by the AMF *Rhizophagus irregularis*. Bar = 1mm.

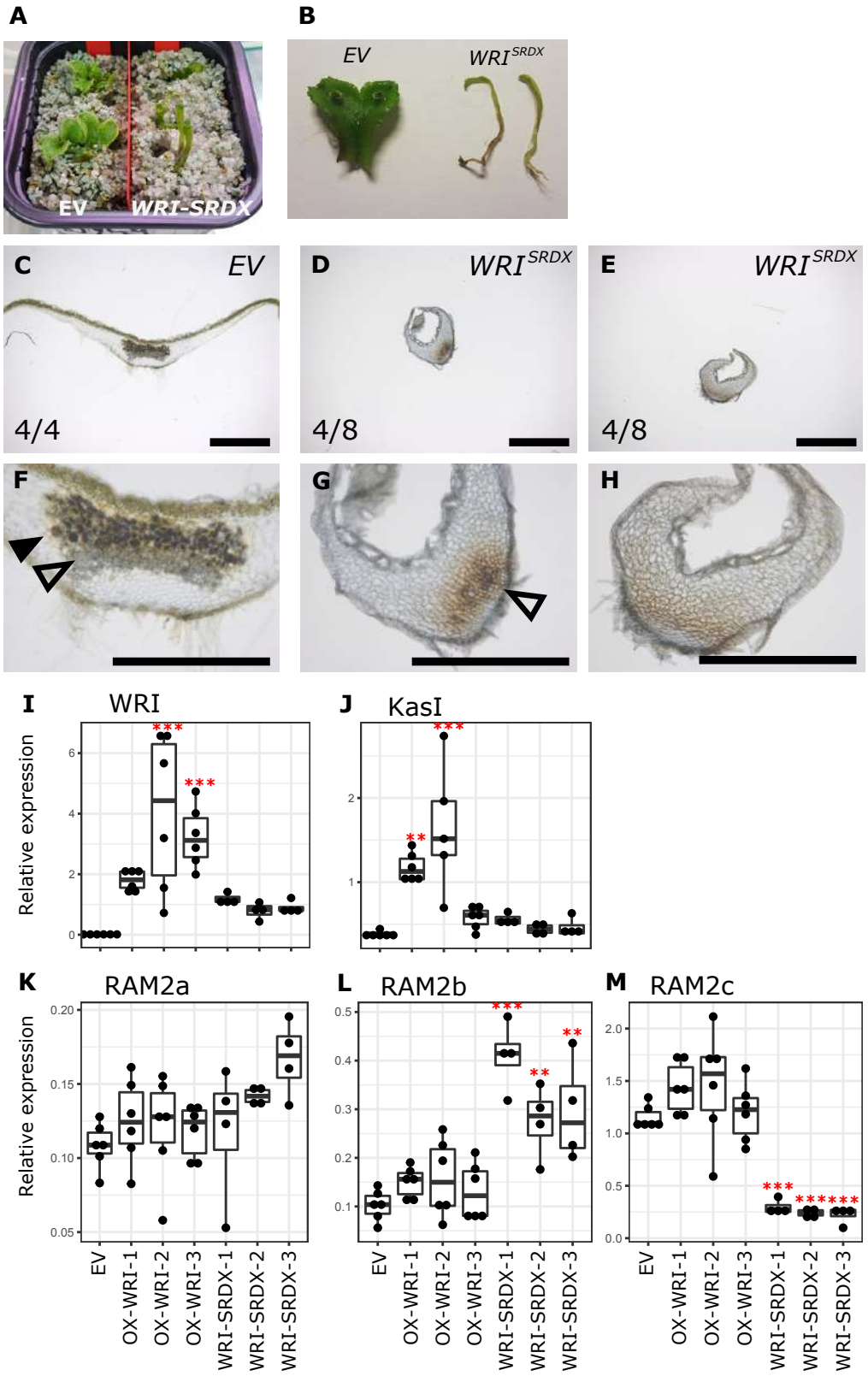

Fig. S12. Dominant negative form of *MpaWRI* causes both developmental and symbiosis defects. (A-B) Expression of *HA:WRI*^{SRDX} causes severe developmental defects. (C-H) Empty vector control (C, F) and *pIndHA:WRI*^{SRDX} (D, E, G, H) ink colored transversal sections. Control plants show colonization of the central part of the thalli with an arbuscule containing zone (black triangle) and a zone colonized with intracellular hyphae (open triangle). In *pIndHA:WRI*^{SRDX}, 4/8 thalli show intracellular colonization without arbuscules and 4/8 are not colonized. Scale bar = 1 mm. (I-M) Relative expression of *WRI* (I), *KasI* (J), *RAM2a* (K), *RAM2b* (L), *RAM2c* (M) in empty vector control, *HA:WRI*^{ox} and *HA:WRI*^{SRDX}.(Anova, post-hoc Tukey p-value <0.05:*, <0.01:**, <0.001: ***)

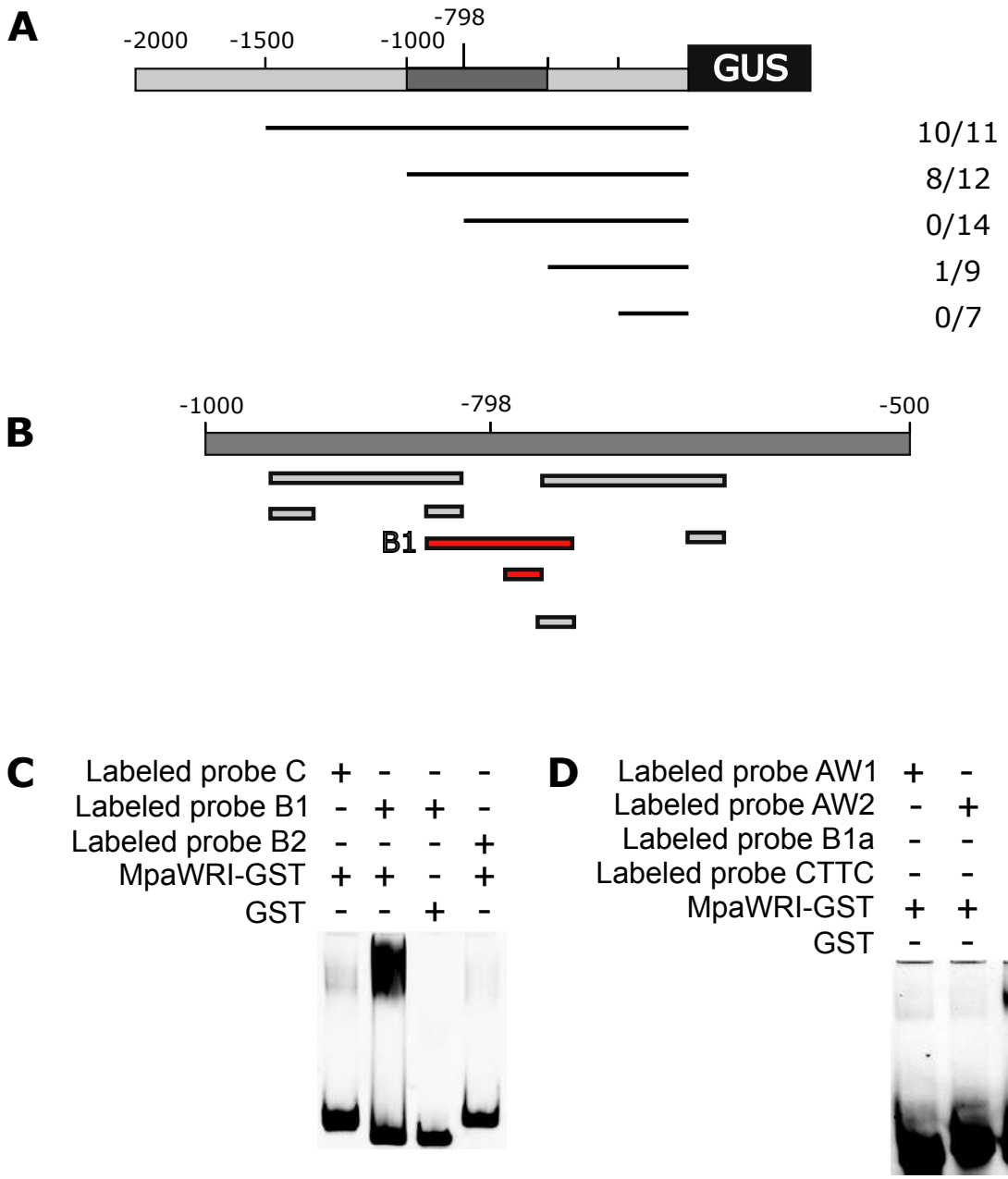

Fig. S13. Regulation of *MpaRAM2* by MpaWRI. (**A**) Promoter bashing analysis of the *MpaRAM2c* promoter. Number of thalli with GUS signal at the arbuscule containing zone are indicated. (**B**) Position of the EMSA probes on the -500 -1000bp zone of the pMpaRAM2c promoter. In red the probes B1 and B1A bound by GST-MpaWRI. (**C-D**) EMSA using GST-MpaWRI show no binding on canonical AW box (CNTNGNNNNNNNCG) but define an alternative motif.

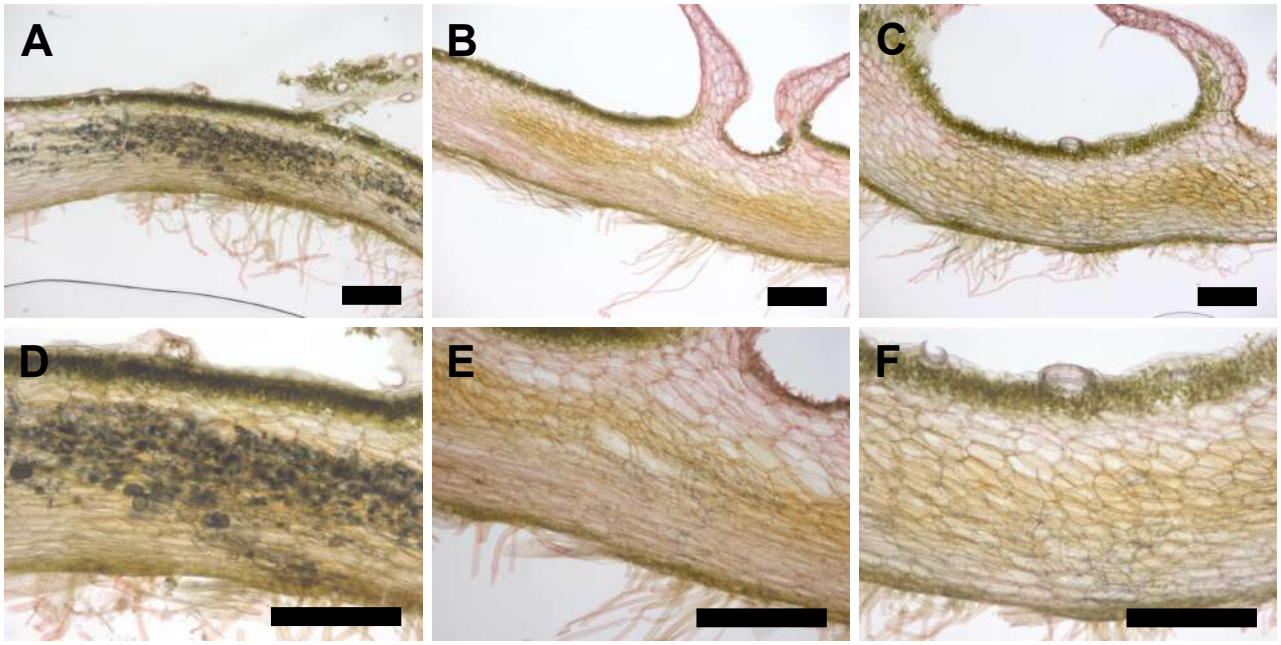

Fig. S14. *Marchantia paleacea ssp. diptera wri* is impaired in arbuscule development. Ink colored longitudinal sections of wild type (A, D) or *wri-8* (B, C, E, F). Bar = 0.5 mm.

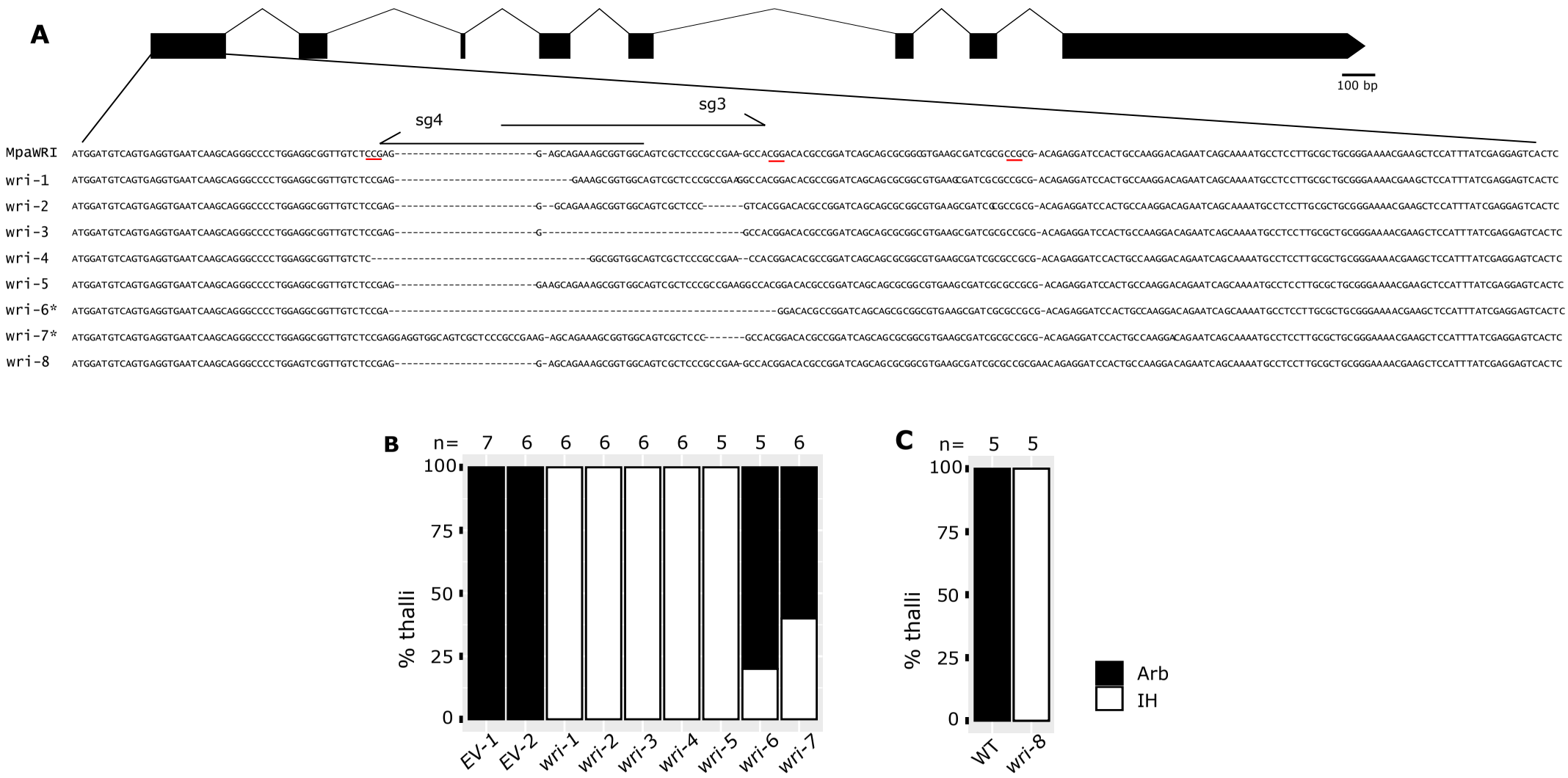

Fig. S15. *MpaWRI* CRIPSR/Cas9 mutants. (**A**) Alignment of *wri-1* to *7* alleles of *Marchantia paleacea ssp. paleacea* and *wri-8* of *Marchantia paleacea ssp. diptera*. Protospacer Adjacent Motif (PAM) are underlined in red. * indicates alleles were edition of the gene did not change the reading frame. (**B-C**) Proportion of thalli showing arbuscules in the *wri-1* to *-7* (**B**) and *wri-8* (**C**) alleles.

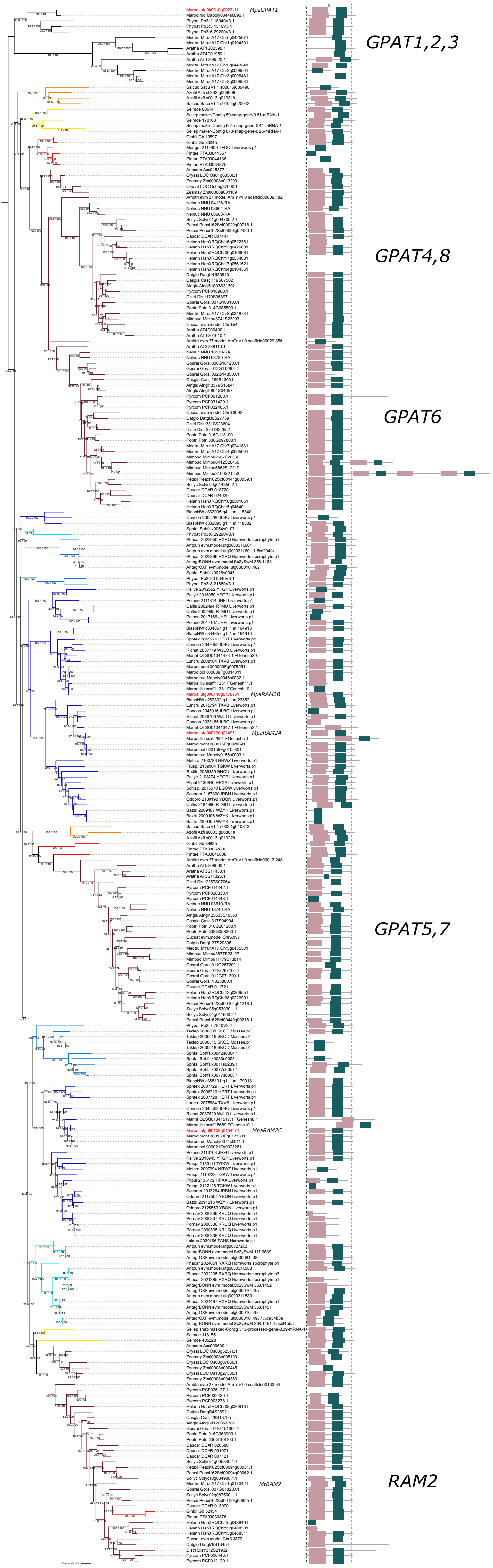

Fig. S16: Maximum likelihood tree of *GPAT4* to *GPAT8* genes (model: SYM+R8; log-likelihood: -195642.6059). Tree is rooted on the representative sequences from GPAT1, 2, 3 clades. GPAT functional domains, HAD (PF12710.8) and Acyltransferase (PF01553.22) are represented by pink and green boxes respectively. Their size and position are proportional to the domain and protein length respectively. Branches are colored according to the main plant lineage as described at the bottom of the tree.

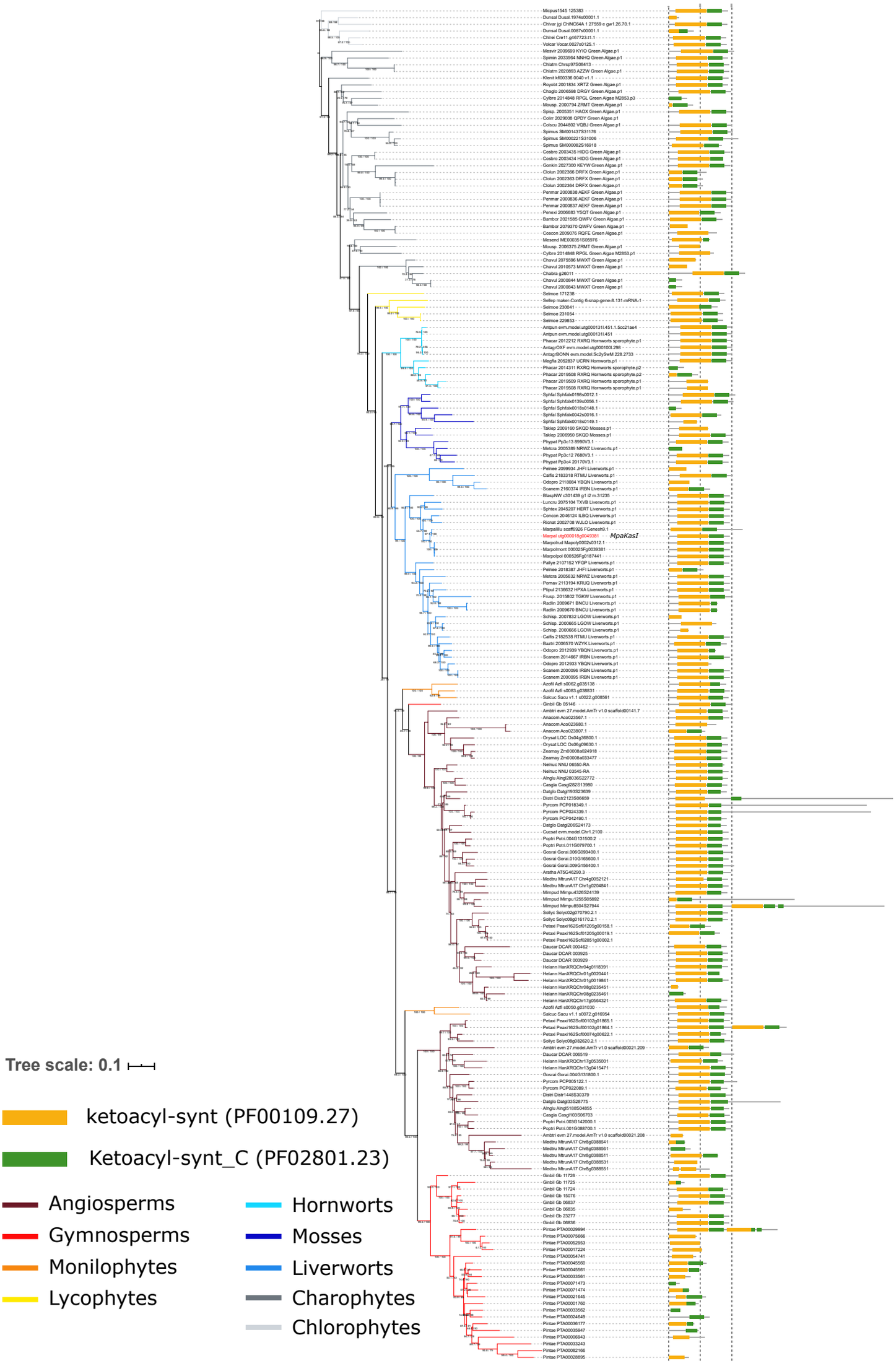

Fig. S17: Maximum likelihood tree of *KASI* genes (model: SYM+R6; log-likelihood: -103534.5453). Tree is rooted on the Chlorophyta lineage. KAS functional domains, PF00109.27 and PF02801.23 are represented by orange and green boxes respectively. Their size and position is proportional to the domain and protein length respectively. Branches are colored according to the main plant lineage as described at the bottom of the tree.

Fig. S18: Maximum likelihood tree of *Fat* genes (model: SYM+R7; log-likelihood: -235670.6868). Tree is rooted on the Chlorophyta lineage. Fat functional domains PF001643.18 and PF12590.9 are represented by blue and violet boxes respectively. Their size and position is proportional to domain length and protein length respectively. Branches are colored according to the main plant lineage as described at the bottom of the tree.