An ancestral signalling pathway is conserved in intracellular symbioses-forming plant lineages

Guru Radhakrishnan, Jean Keller, Mélanie K Rich, Tatiana Vernié, Duchesse Lacour Mbadinga Mbadinga, Ludovic Cottret, Hélène San Clemente, Cyril Libourel, Jitender Cheema, Anna-Malin Linde, et al.

An ancestral signalling pathway is conserved in plant lineages forming intracellular symbioses

Guru V. Radhakrishnan[1]*, Jean Keller[2]*, Melanie K. Rich[2]*, Tatiana Vernié[2]*, Duchesse L. Mbadinga Mbaginda[2], Nicolas Vigneron[2], Ludovic Cottret[3], Hélène San Clemente[1], Cyril Libourel[2], Jitender Cheema[1], Anna-Malin Linde[4], D. Magnus Eklund[4], Shifeng Cheng[5], Gane KS Wong[6,7,8], Ulf Lagercrantz[4], Fay-Wei Li[9,10], Giles E. D. Oldroyd[1,11§], Pierre-Marc Delaux[2§]

[1] John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom

[2] LRSV, Université de Toulouse, CNRS, UPS, 31326 Castanet-Tolosan, France

[3] LIPM, Université de Toulouse, INRA, CNRS, 31326 Castanet-Tolosan, France.

[4] Plant Ecology and Evolution, Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, 75236 Uppsala, Sweden

[5] Agricultural Genome Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518124, China

[6] BGI-Shenzhen, Shenzhen 518083, China.

[7] Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada.

[8] Department of Medicine, University of Alberta, Edmonton, AB T6G 2E, Canada.

[9] Boyce Thompson Institute, Ithaca, New York, USA.

[10] Plant Biology Section, Cornell University, New York, USA.

[11] Sainsbury Laboratory, University of Cambridge, Bateman Street, Cambridge CB2 1LR, UK

* These authors contributed equally.

[§] emails: pierre-marc.delaux@lrsv.ups-tlse.fr & gedo2@cam.ac.uk

ABSTRACT

Plants are the foundation of terrestrial ecosystems and their colonization of land was likely facilitated by mutualistic associations with arbuscular mycorrhizal fungi. Following that founding event, plant diversification has led to the emergence of a tremendous diversity of mutualistic symbioses with microorganisms, ranging from extracellular associations to the most intimate intracellular associations, where fungal or bacterial symbionts are hosted inside plant cells. Through analysis of 271 transcriptomes and 122 plant genomes, we demonstrate that the common symbiosis signalling pathway controlling the association with arbuscular mycorrhizal fungi and with nitrogen-fixing bacteria specifically co-evolved with intracellular endosymbioses, including ericoid and orchid mycorrhizae in angiosperms and ericoid-like associations of bryophytes. In contrast, species forming exclusively extracellular symbioses like ectomycorrhizae or associations with cyanobacteria have lost this signalling pathway. This work unifies intracellular symbioses, revealing conservation in their evolution across 450 million years of plant diversification.

Introductory paragraph

Since they colonized land 450 million years ago, plants have been the foundation of most terrestrial ecosystems[1]. Such successful colonization occurred only once in the plant kingdom and it has been proposed that the symbiotic association formed with arbuscular mycorrhizal fungi supported that transition[2,3]. Following that founding event, plant diversification was accompanied by the emergence of alternative or additional symbionts[4]. Among alternative symbioses, the association between orchids and Ericales with both ascomycetes and basidiomycetes are two endosymbioses with specific intracellular structures in two plant lineages that lost the ability to form the Arbuscular Mycorrhizal Symbiosis (AMS)[5]. As such, orchid mycorrhiza and ericoid-mycorrhiza represent two clear symbiosis switches, whereby intracellular associations are sustained, but the nature of the symbionts are radically different. Similarly, within the liverworts, the Jungermanniales engage in ericoid-like endosymbioses but not AM symbiosis and represent another symbiont switch that occurred during plant evolution[6]. Other symbioses can occur simultaneously with AMS, for example the root nodule symbiosis, an association with nitrogen-fixing bacteria that evolved in the last common ancestor of Fabales, Fagales, Cucurbitales and Rosales[7]. Another example is ectomycorrhizae, an extracellular symbiosis found in several gymnosperm and angiosperm lineages: in some lineages both AMS and ectomycorrhizae have been retained; while other lineages have switched from AMS to ectomycorrhizae[8]. Finally, associations with cyanobacteria, which occur only in the intercellular spaces of the plant tissue, can be found in diverse species within the embryophytes, in hornworts, liverworts, ferns, gymnosperms and angiosperms[9]. Despite the improved nutrient acquisition afforded to plants by these different types of mutualistic symbioses, entire plant lineages have completely lost the symbiotic state, a phenomenon known as mutualism abandonment[4].

Our understanding of the molecular mechanisms governing the establishment and function of these symbioses comes from forward and reverse genetics conducted in legumes and a few other angiosperms[10] and restricted to AMS and the root-nodule symbiosis[7,11]. These detailed studies in selected plant species have allowed phylogenetic analyses to more precisely link the symbiotic genes with either AMS or root-nodule symbiosis. Indeed, the loss of AMS or the root-nodule symbiosis correlates with the loss of many genes known to be involved in these associations[7,11]. The gene losses are thought to be the result of relaxed selection following loss of the trait, resulting in co-elimination, that specifically targets genes only required for the lost trait[12]. Co-elimination can be tracked at the genome-wide level using comparative phylogenomic approaches on species with contrasting retention of the trait of interest[13,14]. Such approaches led for instance to the discovery of genes associated with small RNA biosynthesis and signalling[13] or cilia function[15]. Applied to the AMS, comparative

phylogenomics in angiosperms identified a set of more than 100 genes that were lost in a convergent manner in lineages that lost the AMS[16–18]. All classes of functions essential for AMS were detected among these genes, including the initial signalling pathway, essential for the host plant to activate its symbiotic program, and genes involved in the transfer of lipids from the host plant to the AM fungi. Since their identification by phylogenomics, novel candidates were validated for their involvement in the AM symbiosis through reverse genetic analyses in legumes[19–21].

Targeted phylogenetic analyses have identified multiple symbiotic genes in the transcriptomes of bryophytes, but study on the overall molecular conservation of symbiotic mechanisms in land plants are lacking[22–24]. Similarly, knowledge on the plant molecular mechanisms behind the diverse array of mutualistic associations, either intracellular or extracellular, are poorly understood[10]. Here, we demonstrate through analysis of a comprehensive set of plant genomes and transcriptomes, the loss and conservation of symbiotic genes associated with the evolution of diverse mutualistic symbioses in plants.

RESULTS AND DISCUSSION

A database covering the diversity of plant lineages and symbiotic associations. Genomic and transcriptomic data are scattered between public repositories, specialized databases and personal websites. To facilitate large scale phylogenetic analysis, we compiled resources for species covering the broad diversity of plants and symbiotic status in a centralized database, SymDB (www.polebio.lrsv.ups-tlse.fr/symdb/). This sampling of available resources covers lineages forming most of the known mutualistic associations in plants (Supplementary Table 1), including the AM symbiosis, root nodule symbiosis, ectomycorrhizae, orchid mycorrhiza, cyanobacterial associations in hornworts and ferns, and ericoid-like symbioses in liverworts. SymDB also includes genomes of lineages that have abandoned mutualism in the angiosperms, gymnosperms, monilophytes, and bryophytes. To enrich this sampling, we generated two additional datasets: an *in depth* transcriptome of the liverwort *Blasia pusilla* that associates with cyanobacteria and since no genome of an AM-host was available for the bryophytes, we *de novo* sequenced the genome of the complex thalloid liverwort *Marchantia paleacea* which specifically associates with AM fungi[25]. The obtained assembly was of similar size and completeness to the *Marchantia polymorpha* TAK1 genome[26] (Table 1). These new datasets augmented the SymDB database encompassing a total of 125 genomes and 271 transcriptomes that provided broad coverage of mutualistic symbioses in plants.

Mutualism abandonment leads to gene loss, positive selection or pseudogenization of symbiosis genes. Previous studies have demonstrated that loss of AMS in six angiosperm

lineages is associated with the convergent loss of many genes[16,17]. SymDB contains species from across the entire land plant lineage that have lost AMS and thus provided us with a platform to assess co-elimination of genes associated with the abandonment of mutualism throughout the plant kingdom. We generated phylogenetic trees for all the genes identified previously as being lost in angiosperms with loss of AMS (Supplementary Fig. 1-32 and Supplementary Table 2). Among these gene phylogenies, those missing from most lineages that have abandoned mutualism were selected. Six genes, *SymRK*, *CCaMK*, *CYCLOPS,* the GRAS transcription factor RAD1 and two half ABCG transporters STR and STR2 were consistently lost in non-mutualistic lineages in angiosperms, gymnosperms, ferns and Bryophytes (Fig. 1; Supplementary Fig. 1,3,13–15). Very few exceptions to this trend were found (Fig. 1; Supplementary Fig. 13–14), for instance the presence of *CCaMK* in the aquatic angiosperm *Nelumbo nucifera,* that was previously reported in Bravo *et al*[16]. However, further analysis of this locus revealed a deletion in the kinase domain leading to a likely non-functional pseudogene (Supplementary Fig. 33). The same deletion was present in two different ecotypes and three independent genome assemblies (Supplementary Fig. 33). The second significant exception was in mosses, where *CCaMK* and *CYCLOPS* were present despite the documented loss of AMS in this lineage (Fig. 1; Supplementary Fig. 13–14). Previously, it was proposed that the selection acting on both genes was relaxed in the branch following the divergence of the only mycorrhizae-forming moss (*Takakia*) and other moss species[24]. Using two independent approaches (RELAX and PAML) we confirmed this initial result and identified sites under positive selection (Supplementary Fig. 34–35; Supplementary Table 3), suggesting the neofunctionalization of these two genes. From our analysis we also detected in three species that are thought to be non-mutualistic the presence of STR2 (in a single fern species) and RAD1 (in two liverworts) (Fig. 2b,c). The presence of these genes may reflect additional cases of species-specific neofunctionalization or may be the result of misassignment of the symbiotic state[27].

Among the species that have abandoned mutualism, *M. polymorpha* is a particularly intriguing case, since the non-mutualistic *M. polymorpha* and the mutualistic *M. paleacea* belong to recently diverged lineages[28]. *M. polymorpha* represents the most recent loss of mutualism of which we are aware and hence this species may allow us to witness the process of co-elimination of symbiosis genes with the loss of mutualism. For this reason we chose to sequence the genome of the symbiotic *M. paleacea* to allow a detailed comparison between symbiotic and non-symbiotic liverwort species. Using microsynteny we identified potential remnants of symbiotic genes in *M. polymorpha* ssp. *ruderalis* TAK1, pseudogenes for *SymRK*, *CCaMK*, *CYCLOPS* and *RAD1* existed in genomic blocks syntenic with *M. paleacea* while *STR* and *STR2* were completely absent (Fig. 3). These pseudogenes have accumulated point

mutations, deletions and insertions (Fig. 3) and their presence supports a recent abandonment of mutualism in *M. polymorpha*. To better position the timing of this abandonment, we collected 35 *M. polymorpha* ssp. *ruderalis* accessions in Europe (Supplementary Table 4) and sequenced *CYCLOPS* and *CCaMK*. All accessions harboured pseudogenes at these two loci, confirming the fixation of these null alleles in the subspecies *ruderalis* (Fig. 3 and Supplementary Figure 36). Besides *M. polymorpha* ssp. *ruderalis*, two other *M. polymorpha* subspecies have been reported, ssp. *polymorpha* and ssp. *montivagans,* that are sister to *M. paleacea*. We phenotyped these two subspecies in controlled conditions, and confirmed that the loss of the AM symbiosis occurred before the radiation of the three *M. polymorpha* subspecies approximately 5 million years ago[28] (Fig. 3). We sequenced high-quality genomes of *M. polymorpha* ssp. *montivagans* and *polymorpha* and searched for the presence of the six aforementioned genes. As for *M. polymorpha* ssp. *ruderalis* all six genes were pseudogenized or missing in these two novel assemblies (Fig. 3). As expected for genes under relaxed selection, the signatures of pseudogenization were different between the three subspecies (Fig. 3).

We conclude that mutualism abandonment leads to the consistent loss, pseudogenization or relaxed selection of at least six symbiosis-specific genes in all surveyed land plant lineages.

Genes specific to AMS in land plants. We have shown consistent loss of six genes with mutualism abandonment, but with the broader array of genomes present we are now able to test whether these genes are lost with mutualism abandonment or specifically with the loss of AMS. Three genes, *RAD1*, *STR* and *STR2*, show a phylogenetic pattern consistent with gene loss specifically associated with the loss of AMS (Fig. 2). Our dataset covers at least 29 convergent losses of AMS, thus representing many independent replications of AMS loss in vascular plants and in bryophytes. We therefore conclude that *RAD1, STR* and *STR2* were specific to AMS in the most recent common ancestor of all land plants. The fact that all three genes are absent from non-AM host lineages indicates a particularly efficient co-elimination of these genes following the loss of AMS, suggesting possible selection against these genes[4,17,27]. We suggest that one potential driver for the loss of AMS is the adaptation to nutrient-rich ecological niches, which are known to inhibit the formation of AMS[29] and thus render the symbiosis redundant. Alternatively, selection against these genes may be driven by the hijacking of the AMS-related pathway by pathogens that would result in positive selection acting against this pathway in the presence of a high pathogen pressure. Although this question is not settled yet, the example of *RAD1,* which has been demonstrated to act as a susceptibility factor to the oomycete pathogen *Phytophthora palmivora* in *Medicago truncatula*[30]*,* provides support for this hypothesis. *RAD1* encodes a transcription factor in the GRAS family and *rad1* mutants display reduced colonization by AM fungi, defective

arbuscules, the interface for nutrient exchange formed by both partners inside the plant cells and reduced expression of *STR* and *STR2*[19,31]. These two half-ABC transporters are present on the peri-arbuscular membrane, are essential for functional AMS, and have been proposed as involved in the transfer of lipids from the host plant to AM fungi[20,21,32,33]. The specialization of *RAD1*, *STR* and *STR2* to AMS in all plant lineages analysed supports an ancient ancestral origin in land plants for symbiotic lipid transfer to AMS.

The symbiotic signalling pathway is conserved in species with intracellular symbioses. In contrast to *RAD1*, *STR* and *STR2*, the symbiosis signalling genes *CCaMK*, *CYCLOPS* and *SymRK* are not absent from all species that have lost AMS. To understand this mixed phylogenetic pattern, we investigated their conservation across species with diverse symbioses. *CCaMK*, *CYCLOPS* and *SymRK* were absent from seven genomes and fourteen transcriptomes of Pinaceae, that form ectomycorrhizae, but not AMS. None of these genes were detected in the genome of the fern *Azolla filiculoides* or in the transcriptome of the liverwort *B. pusilla,* that have independently evolved associations with nitrogen-fixing cyanobacteria, but lost AMS[9] (Fig. 2, Supplementary Fig. 13–15). During ectomycorrhizae, the symbiotic fungi colonize the intercellular space between epidermal cells and the first layer of cortical cells[5]. Similarly, in both *A. filiculides* and *B. pusilla,* nitrogen-fixing cyanobacteria are hosted in specific glands, but outside plant cells[34,35]. Therefore, all the lineages in our sampling that host fungal or bacterial symbionts exclusively outside their cells did not retain *SymRK, CCaMK* and *CYCLOPS* suggesting their dispensability for extracellular symbiosis. Confirming this, knock-down analysis of *CCaMK* in poplar, which forms both AMS and ectomycorrhizae, resulted in only a quantitative decrease in ectomycorrhizae, while AMS was completely aborted[36].

All other lineages that switched from AMS to other types of mutualistic symbioses retained the three signalling genes *SymRK, CCaMK and CYCLOPS* (Fig. 2, Supplementary Fig. 13–15). These lineages are scattered throughout the land plant phylogeny, thus excluding the hypothesis of a lineage-specific retention of these genes. The three genes were found in the genomes of three Orchidaceae, *Apostasia shenzhenica*, *Dendrobium catenatum* and *Phalaenopsis equestris* and in the transcriptome of *Bletilla striata*[37], that form Orchid Mycorrhizae with Basidiomycetes that develop intracellular pelotons[5]. All three genes were also detected in the transcriptomes of liverworts from the Jungermaniales order *Scapania nemorosa* (3/3), *Calypogeia fissa* (1/3), *Odontoschisma prostratum* (1/3), *Bazzania trilobata* (2/3) and *Schistochila sp.* (1/3), which have switched from AMS to diverse Ericoid-like associations with Basidiomycetes and Ascomycetes that form intracellular coils[5]. Furthermore, these genes are also conserved in the legume genus *Lupinus,* which associates with nitrogen-fixing rhizobia forming intracellular symbiosomes in root nodules, but have lost AMS[38].

A unifying feature of these species that have preserved the symbiosis signalling pathway, but lost AMS, is their ability to engage in alternative intracellular mutualistic symbioses and we therefore suggest that these genes may be conserved specifically with intracellular symbioses throughout the plant kingdom. To test this hypothesis, we added to our initial analysis the only known intracellular mutualistic symbiosis not covered in our sampling: ericoid mycorrhizae, which evolved before the radiation of the angiosperm family Ericaceae. In ericoid-mycorrhiza, ascomycetes colonize epidermal root cells and develop intracellular hyphal complexes[5]. We collected available transcriptomic data from six species, assembled and annotated them, and specifically searched for the presence of *SymRK, CCaMK*, *CYCLOPS,* as well as *STR*, *STR2* and *RAD1*. Congruent with the loss of AMS, neither *RAD1*, *STR* nor STR2 were detected (Supplementary Fig. 37), but *SymRK, CCaMK* and *CYCLOPS* were all recovered from the transcriptome of *Rhododendron fortunei* roots, but not in the other five transcriptomes derived from leaves or stems (Fig. 2, Supplementary Fig. 37).

The receptor-like kinase SYMRK, the Calcium and calmodulin dependent protein kinase CCaMK and the transcription factor CYCLOPS are known components of the common symbiosis signalling pathway and contribute successive steps in the signalling processes triggered by AM fungi and nitrogen-fixing bacteria[2,39]. Genetic analysis of *SymRK*, *CCaMK* and *CYCLOPS* have been conducted in multiple angiosperms, including dicots from the Fabaceae, Casuarinaceae, Fagaceae, Rosaceae or Solanaceae families as well as monocots such as rice[40–43]. In all these species, defects in any of these three genes resulted in aborted, or strongly attenuated, intracellular infection by AM fungi. In addition, knock-out or knock-down in root-nodule symbiosis-forming species resulted in impaired intracellular infection by nitrogen-fixing bacteria[2,39]. Conversely, *CCaMK* knock-down in the Fabaceae *Sesbania rostrata* did not impact extracellular infection of cortical cells by nitrogen-fixing rhizobia[44]. Together with this genetic evidence, our results demonstrate that *SymRK*, *CCaMK* and *CYCLOPS* specifically occur in species accommodating intracellular symbionts, defining a universal signalling pathway for intracellular mutualistic symbioses in plants.

Conservation of CCaMK and CYCLOPS biochemical properties in land plants. We propose that the symbiosis signalling pathway has been co-opted for all intracellular endosymbioses in land plants and this would imply conservation of the biochemical properties of the corresponding proteins over the 450 million years of land plant evolution. To test this hypothesis, we cloned *CCaMK* from three dicots forming AMS or both AMS and root-nodule symbiosis, from two monocots forming only the AMS and from the liverwort *M. paleacea* that forms AMS in the absence of roots. Two assays were used to assess the conservation of the biochemical properties of these CCaMK orthologs. First, truncated versions that only contain the kinase domain of CCaMK (CCaMK-K) were cloned under control of a constitutive promoter.

If functional, these constructs are expected to induce the expression of root-nodule symbiosis reporter genes such as *ENOD11*[45] when overexpressed in the Fabale *Medicago truncatula* roots in absence of symbiotic bacteria, as does the *M. truncatula CCaMK-K* construct[45]. All the constructs were introduced in a *M. truncatula pENOD11:GUS* background and GUS activity monitored in the absence of symbiotic bacteria. CCaMK-K from every tested species resulted in the spontaneous activation of the *pENOD11:GUS* reporter (Fig 5). As a second test of the conservation of CCaMK, trans-complementation assays of a *M. truncatula ccamk* (*dmi3*) mutant were complemented with *CCaMK* orthologs from the above mentioned species. In the presence of symbiotic bacteria, all of the CCaMK orthologs were able to restore nodule formation and intracellular infection in the *ccamk* mutant (Fig. 4 and Supplementary Table 5).

In legumes, CCaMK phosphorylates CYCLOPS. Phosphorylated CYCLOPS then binds to the promoter and activates the transcription of downstream genes[46]. To determine whether the CCaMK-CYCLOPS module itself is biochemically conserved across land plants, we conducted trans-complementation assays of a *M. truncatula cyclops* (*ipd3*) mutant with *M. paleacea CYCLOPS*. Nodules can be formed in the *M. truncatula cyclops* mutant due to the presence of a functional paralog[47]. In our assay, *M. truncatula cyclops* mutants transformed with the empty vector could develop root nodules, but were mostly uninfected. By contrast, transformation with either *M. truncatula CYCLOPS* or *M. paleacea CYCLOPS* resulted in the formation of many fully infected nodules in most of the transformed *cyclops* roots (Fig. 4).

Altogether, these assays indicate that CYCLOPS and CCaMK orthologs that evolved in different symbiotic (AMS, root-nodule symbiosis, both) and developmental (gametophytes in *M. paleacea*, root sporophytes in angiosperms) contexts have conserved biochemical properties.

Infection-related genes are conserved in angiosperms with intracellular symbioses. For a given gene with dual biological functions, co-elimination is not predicted to occur following the loss of a single trait because of the selection pressure exerted by the other, still present, trait[12]. For instance, DELLA proteins that are involved in AMS and are essential players of gibberellic-acid signalling are retained in all embryophytes[48]. To become sensitive to co-elimination, a given gene must become specific to a single trait. This may occur via the successive losses of the traits or via gene duplication leading to subfunctionalization between the two paralogs[12]. Angiosperm genomes experienced multiple rounds of whole-genome duplications[49] and we hypothesized that, besides the common symbiotic signalling pathway, other genes might be specialized to intracellular symbioses following subfunctionalization. We screened our phylogenies for genes that are retained in angiosperm species forming intracellular symbioses but lost in those that have experienced mutualism abandonment. Six

genes followed that pattern: *KinF*, *EPP1*, *VAPYRIN, LIN/LIN-like, CASTOR* and *SYN* (Fig. 5, Supplementary Fig. 9, 16, 21, 22, 23 and 28). Among them, CASTOR and to some extent EPP1[50] are other components of the common symbiotic signalling pathway, while, VAPYRIN and LIN/LIN-like are directly involved in the formation of a structure required for the intracellular accommodation of rhizobial bacteria[51–53], and also functioning in intracellular accommodation of arbuscular mycorrhizal fungi[54–56]. Similarly, *SYN* has been characterized in *M. truncatula* for its role in the formation of the intracellular structures during both AMS and root-nodule symbiosis[57].

These results demonstrate that, besides the common symbiotic signalling pathway, genes directly involved in the intracellular accommodation of symbionts are exclusively maintained in species that form intracellular mutualistic symbioses in angiosperms, irrespective of the type of symbiont or the plant lineage. Pro-orthologs of these genes are found in species outside the angiosperms, in both symbiotic and non-symbiotic species. Given the overall cellular and molecular conservation observed in AMS processes in land plants, we hypothesize that these genes most likely have an endosymbiotic function in species outside of the angiosperms, but the lack of gene erosion of these genes in non-angiosperms suggests that they have an additional function that ensures their retention. In angiosperms we see loss of these genes concomitant with the loss of intracellular symbioses, suggesting either that their essential function is now redundant in angiosperms or is supported by gene paralogs resulting from the whole genome duplications that predate modern angiosperms.

CONCLUDING REMARKS

Through comprehensive phylogenomics of previously unexplored plant lineages, we demonstrate that three genes are evolutionary linked to the AM symbiosis in all land plants, including two directly involved in the transfer of lipids from the host plant to AM fungi. We propose that the symbiotic transfer of lipids has been essential for the conservation of AMS in land plants. Surprisingly, we demonstrate that genes associated with symbiosis signalling are invariantly conserved in all land plant species possessing intracellular symbionts, implying the repeated recruitment of this signalling pathway, independent of the nature of the intracellular symbiont. Furthermore, we see evidence for conservation of genes associated with the formation of the cellular structure necessary for intracellular accommodation of symbionts, but this correlation is restricted to angiosperms that have duplicated many of these components. Our results provide compelling evidence for the early emergence of genes associated with accommodation of intracellular symbionts, with the onset of AMS in the earliest land plants and then recruitment and retention of these processes during symbiont switches that have occurred on many independent occasions in the 450 million years of land plant evolution (Fig. 6). Our

work also suggests that mutualistic interactions involving extracellular symbionts do not utilise the same molecular machinery as intracellular symbioses, suggesting an alternative evolutionary trajectory for the emergence of ectomycorrhizal and cyanobacterial associations.

Methods

Genome and transcriptome sequencing

Marchantia paleacea

Plant material. *M. paleacea* thalli previously collected in Mexico (Humphreys) were kindly provided by Dr Katie J. Field and Professor David J. Beerling (The University of Sheffield). Gemma from these thalli were collected using a micropipette tip and placed into a micro-centrifuge tube. 5% sodium hypochlorite solution was used to sterilize the gemmae for about 30 seconds followed by rinsing with sterile water 5 times to remove residual sodium hypochlorite solution. The sterilized gemmae were grown on Gamborg's half-strength B5 medium on sterile tissue culture plates under a 16/8-hour day-night cycle at 22° C under fluorescent illumination with a light intensity of 100 µmol/µm^{2}s.

DNA and RNA extraction. Genomic DNA from 8 week old *M. paleacea* thalli were extracted as described previously[58]. RNA extraction was also carried out from 8 week old *M. paleacea* thalli using the RNeasy mini plant kit following the manufacturer's protocols. Fragmentation of RNA and cDNA synthesis were done using kits from New England Biolabs (Ipswich, MA) according to the manufacturer's protocols with minor modifications. Briefly, 1µg of total RNA was used to purify mRNA on Oligo dT coupled to paramagnetic beads (NEBNext Poly(A) mRNA Magnetic Isolation Module). Purified mRNA was fragmented and eluted from the beads in one step by incubation in 2x first strand buffer at 94°C for 7 min, followed by first strand cDNA synthesis using random primed reverse transcription (NEBNext RNA First Strand Synthesis Module), followed by random primed second strand synthesis using an enzyme mixture of DNA PolI, RnaseH and E.coli DNA Ligase (NEBNext Second Strand Synthesis Module).

Genome sequencing. For the *M. paleacea* genome sequencing, short-insert paired-end and long-insert mate-pair libraries were produced. For the paired-end library, the DNA fragmentation was done using the Covaris (Covaris Inc., Woburn, MA). The NEBNext Ultra DNA Library Prep Kit (New England Biolabs, Ipswich, MA) was used for the library preparation, and the bead size selection for 300-400 bp based on the manufacturer's protocol. The library had an average insert size of 336 bp. For the mate-pair library, preparation was done using the Nextera Mate Pair DNA library prep kit (Illumina, San Diego, CA) following the manufacturer's protocol. After enzymatic fragmentation, we used gel size selection for 3-5 kb fragments. The average size that was recovered from the gel was 4311 bp. Sequencing was

carried out on an Illumina HiSeq2500 on 2x100bp Rapid Run mode. The library preparation and sequencing were carried out by GENEWIZ (South Plainfield, NJ).

Genome assembly. Adapter and quality trimming were performed on the paired-end library using Trimmomatic v0.33 using the following parameters (ILLUMINACLIP:TruSeq2-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:12). The trimmed paired-end reads were carried forward for assembling the contigs using multiple assemblers. Scaffolding of the contigs was done using the scaffolder of SOAPdenovo2[59] with the mate-pair library after processing the reads through the NextClip[60] pipeline to only retain predicted genuine long insert mate-pairs. Assembly completeness was measured using the BUSCO[61] plants dataset.

Transcriptome sequencing. cDNA was purified and concentrated on MinElute Columns (Qiagen) and used to construct an Illumina library using the Ovation Rapid DR Multiplex System 1-96 (NuGEN, Redwood City, CA). The library was amplified using MyTaq (Bioline, London, UK) and standard Illumina TruSeq amplification primers. PCR primer and small fragments were removed by Agencourt XP bead purification. The PCR components were removed using an additional purification on Qiagen MinElute Columns. Normalisation was done using Trimmer Kit (Evrogen, Moscow, Russia). The normalized library was re-amplified using MyTaq (Bioline, London, UK) and standard Illumina TruSeq amplification primers. The normalized library was finally size selected on a LMP-Agarose gel, removing fragments smaller than 350Bp and those larger than 600Bp. Sequencing was done on an Illumina MiSeq on 2 x 300bp mode. The RNA extractions, cDNA synthesis library preparation and sequencing were carried out by LGC Genomics GmbH (Berlin, Germany).

Blasia pusilla

Blasia pusilla was originally collected from Windham County, Connecticut, USA, and maintained in Duke University greenhouse. RNA was extracted from plants with symbiotic cyanobacterial colonies using Sigma Spectrum Plant Total RNA kit. Library preparation and sequencing were done by BGI-Shenzhen. A Ribo-Zero rRNA Removal Kit was used to prepare the transcriptome library. In total, 3 libraries were constructed, which were sequenced on the Illumina Platform Hiseq2000, 150bp paired-ends, with insert size 200bp.

Marchantia polymorpha subspecies

Plant material and DNA extraction. Sterilized gemmae from one individual each of *M. polymorpha* ssp. *montivagans* (sample id MpmSA2) and *M. polymorpha* ssp. *polymorpha*

(sample id MppBR5) were grown as for *M. paleacea* and isolated for DNA extraction. DNA was extracted with a modified CTAB protocol[58].

Genome sequencing. DNA were sequenced with Single-molecule real-time (SMRT) sequencing technology developed by Pacific BioSciences on a PacBio Sequel System with Sequel chemistry and sequence depth of 60X[62]..

Genome assembly. The reads were assembled using HGAP 4[63]. Assembly statistics were assessed using QUAST[64] version 4.5.4, BUSCO[61] version 3.0.2 and CEGMA[65] version 2.5.

Supplementary material and method information is available for this paper.

Data availability

All assemblies and gene annotations generated in this project can be found in SymDB (www.polebio.lrsv.ups-tlse.fr/symdb/). Raw sequencing data can be found under NCBI Bioproject PRJNA576233 (*Blasia pusilla*), PRJNA362997 and PRJNA362995 (*Marchantia paleacea* genome and transcriptome respectively) and PRJNA576577 *(Marchantia polymorpha* ssp. *montivagans* and ssp. *polymorpha*).

References

1. Gensel, P. G. *The Emerald Planet: How Plants Changed Earth's History*. *The Quarterly Review of Biology* **83**, (Oxford University Press, 2008).

2. Parniske, M. Arbuscular mycorrhizae: the mother of plant root endosymbioses. *Nat. Rev. Microbiol.* **6**, 763–775 (2008).

3. Delaux, P. M., Séjalon-Delmas, N., Bécard, G. & Ané, J. M. Evolution of the plant-microbe symbiotic 'toolkit'. *Trends Plant Sci.* **18**, 298–304 (2013).

4. Werner, G. D. A. *et al.* Symbiont switching and alternative resource acquisition strategies drive mutualism breakdown. *Proc. Natl. Acad. Sci.* **115**, 5229–5234 (2018).

5. Smith, S. & Read, D. *Mycorrhizal Symbiosis*. (Academic press, 2008).

6. Kottke, I. *et al.* Heterobasidiomycetes form symbiotic associations with hepatics: Jungermanniales have sebacinoid mycobionts while Aneura pinguis (Metzgeriales) is associated with a Tulasnella species. *Mycol. Res.* (2003). doi:10.1017/S0953756203008141

7. Griesmann, M. *et al.* Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science.* **361**, eaat1743 (2018).

8. Wang, B. & Qiu, Y. L. Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* **16**, 299–363 (2006).

9. Delaux, P. M., Radhakrishnan, G. & Oldroyd, G. Tracing the evolutionary path to nitrogen-fixing crops. *Curr. Opin. Plant Biol.* **26**, 95–99 (2015).

10. Martin, F. M., Uroz, S. & Barker, D. G. Ancestral alliances: Plant mutualistic symbioses with fungi and bacteria. *Science* (2017). doi:10.1126/science.aad4501

11. van Velzen, R. *et al.* Comparative genomics of the nonlegume Parasponia reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proc. Natl. Acad. Sci.* 201721395 (2018). doi:10.1073/pnas.1721395115

12. Albalat, R. & Cañestro, C. Evolution by gene loss. *Nat. Rev. Genet.* **17**, 379–391 (2016).

13. Tabach, Y. *et al.* Identification of small RNA pathway genes using patterns of phylogenetic conservation and divergence. *Nature* **493**, 694–698 (2013).

14. Delaux, P. M. Comparative phylogenomics of symbiotic associations. *New Phytol.* **213**, 89–94 (2017).

15. Dey, G., Jaimovich, A., Collins, S. R., Seki, A. & Meyer, T. Systematic Discovery of Human Gene Function and Principles of Modular Organization through Phylogenetic Profiling. *Cell Rep.* **10**, 993–1006 (2015).

16. Bravo, A., York, T., Pumplin, N., Mueller, L. A. & Harrison, M. J. Genes conserved for arbuscular mycorrhizal symbiosis identified through phylogenomics. *Nat. Plants* **2**, (2016).

17. Delaux, P. M. *et al.* Comparative Phylogenomics Uncovers the Impact of Symbiotic Associations on Host Genome Evolution. *PLoS Genet.* **10**, (2014).

18. Favre, P. *et al.* A novel bioinformatics pipeline to discover genes related to arbuscular mycorrhizal symbiosis based on their evolutionary conservation pattern among higher plants. *BMC Plant Biol.* **14**, 333 (2014).

19. Xue, L. *et al.* Network of GRAS Transcription Factors Involved in the Control of Arbuscule Development in *Lotus japonicus*. *Plant Physiol.* **167**, 854–871 (2015).

20. Keymer, A. *et al.* Lipid transfer from plants to arbuscular mycorrhiza fungi. *Elife* **6**, (2017).

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

22. Grosche, C., Genau, A. C. & Rensing, S. A. Evolution of the symbiosis-specific gras regulatory network in bryophytes. *Front. Plant Sci.* (2018). doi:10.3389/fpls.2018.01621

23. Delaux, P.-M. *et al.* Algal ancestor of land plants was preadapted for symbiosis. *Proc. Natl. Acad. Sci.* **112**, 13390–13395 (2015).

24. Wang, B. *et al.* Presence of three mycorrhizal genes in the common ancestor of land plants suggests a key role of mycorrhizas in the colonization of land by plants. *New Phytol.* **186**, 514–525 (2010).

25. Humphreys, C. P. *et al.* Mutualistic mycorrhiza-like symbiosis in the most ancient group of land plants. *Nat. Commun.* **1**, 103 (2010).

26. Bowman, J. L. *et al.* Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome. *Cell* **171**, 287-304.e15 (2017).

27. Brundrett, M. & Tedersoo, L. Misdiagnosis of mycorrhizas and inappropriate recycling of data can lead to false conclusions. *New Phytologist* (2019). doi:10.1111/nph.15440

28. Villarreal A., J. C., Crandall-Stotler, B. J., Hart, M. L., Long, D. G. & Forrest, L. L. Divergence times and the evolution of morphological complexity in an early land plant lineage (Marchantiopsida) with a slow molecular rate. *New Phytol.* (2016). doi:10.1111/nph.13716

29. Read, D. J. & Perez-Moreno, J. Mycorrhizas and nutrient cycling in ecosystems - A journey towards relevance? *New Phytologist* (2003). doi:10.1046/j.1469-8137.2003.00704.x

30. Rey, T. *et al.* The Medicago truncatula GRAS protein RAD1 supports arbuscular mycorrhiza symbiosis and *Phytophthora palmivora* susceptibility. *J. Exp. Bot.* **68**, 5871–5881 (2017).

31. Park, H.-J., Floss, D. S., Levesque-Tremblay, V., Bravo, A. & Harrison, M. J. Hyphal branching during arbuscule development requires RAM1. *Plant Physiol.* (2015). doi:10.1104/pp.15.01155

32. Luginbuehl, L. H. *et al.* Fatty acids in arbuscular mycorrhizal fungi are synthesized by

the host plant. *Science.* **356**, 1175–1178 (2017).

33. Jiang, Y. *et al.* Plants transfer lipids to sustain colonization by mutualistic mycorrhizal and parasitic fungi. *Science.* **356**, 1172–1175 (2017).

34. Adams, D. G. *Symbiotic interactions*. *Ecology of cyanobacteria: their diversity in time and space* (Kluwer Academic Publishers, 2000). doi:10.1016/0303-2647(92)90025-T

35. Li, F. W. *et al.* Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat. Plants* (2018). doi:10.1038/s41477-018-0188-8

36. Cope, K. R. *et al.* The Ectomycorrhizal Fungus Laccaria bicolor Produces Lipochitooligosaccharides and Uses the Common Symbiosis Pathway to Colonize Populus Roots. *Plant Cell* (2019). doi:10.1105/tpc.18.00676

37. Miura, C. *et al.* The mycoheterotrophic symbiosis between orchids and mycorrhizal fungi possesses major components shared with mutualistic plant-mycorrhizal symbioses. *Mol. Plant-Microbe Interact.* (2018). doi:10.1094/MPMI-01-18-0029-R

38. Oba, H., Tawaray, K. & Wagatsuma, T. Arbuscular mycorrhizal colonization in Lupinus and related genera. *Soil Sci. Plant Nutr.* **47**, 685–694 (2001).

39. Oldroyd, G. E. D. Speak, friend, and enter: Signalling systems that promote beneficial symbiotic associations in plants. *Nature Reviews Microbiology* (2013). doi:10.1038/nrmicro2990

40. Gutjahr, C. *et al.* Arbuscular Mycorrhiza-Specific Signaling in Rice Transcends the Common Symbiosis Signaling Pathway. *Plant Cell Online* **20**, 2989–3005 (2008).

41. Gherbi, H. *et al.* SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and Frankiabacteria. *Proc. Natl. Acad. Sci.* **105**, 4928–4932 (2008).

42. Svistoonoff, S. *et al.* The Independent Acquisition of Plant Root Nitrogen-Fixing Symbiosis in Fabids Recruited the Same Genetic Pathway for Nodule Organogenesis. *PLoS One* **8**, e64515 (2013).

43. Buendia, L., Wang, T., Girardin, A. & Lefebvre, B. The LysM receptor-like kinase SlLYK10 regulates the arbuscular mycorrhizal symbiosis in tomato. *New Phytol.* **210**, 184–195 (2016).

44. Capoen, W. *et al.* Calcium spiking patterns and the role of the calcium/calmodulin-

dependent kinase CCaMK in lateral root base nodulation of sesbania rostrata. *Plant Cell* (2009). doi:10.1105/tpc.109.066233

45. Gleason, C. *et al.* Nodulation independent of rhizobia induced by a calcium-activated kinase lacking autoinhibition. *Nature* **441**, 1149–1152 (2006).

46. Singh, S., Katzer, K., Lambert, J., Cerri, M. & Parniske, M. CYCLOPS, A DNA-binding transcriptional activator, orchestrates symbiotic root nodule development. *Cell Host Microbe* **15**, 139–152 (2014).

47. Jin, Y. *et al.* IPD3 and IPD3L function redundantly in rhizobial and mycorrhizal symbioses. *Front. Plant Sci.* (2018). doi:10.3389/fpls.2018.00267

48. Yasumura, Y., Crumpton-Taylor, M., Fuentes, S. & Harberd, N. P. Step-by-Step Acquisition of the Gibberellin-DELLA Growth-Regulatory Mechanism during Land-Plant Evolution. *Curr. Biol.* **17**, 1225–1230 (2007).

49. Soltis, P. S. & Soltis, D. E. Ancient WGD events as drivers of key innovations in angiosperms. *Current Opinion in Plant Biology* (2016). doi:10.1016/j.pbi.2016.03.015

50. Valdés-López, O. *et al.* A Novel Positive Regulator of the Early Stages of Root Nodule Symbiosis Identified by Phosphoproteomics. *Plant Cell Physiol.* **60**, 575-586 (2019) doi: 10.1093/pcp/pcy228.

51. Murray, J. D. *et al.* Vapyrin, a gene essential for intracellular progression of arbuscular mycorrhizal symbiosis, is also essential for infection by rhizobia in the nodule symbiosis of Medicago truncatula. *Plant J.* **65**, 244–252 (2011).

52. Imaizumi-Anraku, H. *et al.* Plastid proteins crucial for symbiotic fungal and bacterial entry into plant roots. *Nature* **433**, 527–531 (2005).

53. Liu, C. W. *et al.* A protein complex required for polar growth of rhizobial infection threads. *Nat. Commun.* (2019). doi:10.1038/s41467-019-10029-y

54. Pumplin, N. *et al.* Medicago truncatula Vapyrin is a novel protein required for arbuscular mycorrhizal symbiosis. *Plant J.* **61**, 482–94 (2010).

55. Feddermann, N. *et al.* The PAM1 gene of petunia, required for intracellular accommodation and morphogenesis of arbuscular mycorrhizal fungi, encodes a homologue of VAPYRIN. *Plant J.* **64**, 470–481 (2010).

56. Takeda, N., Tsuzuki, S., Suzaki, T., Parniske, M. & Kawaguchi, M. CERBERUS and

NSP1 of lotus japonicus are common symbiosis genes that modulate arbuscular mycorrhiza development. *Plant Cell Physiol.* **54**, 1711–1723 (2013).

57. Huisman, R. *et al.* A symbiosis-dedicated SYNTAXIN OF PLANTS 13II isoform controls the formation of a stable host-microbe interface in symbiosis. *New Phytol.* (2016). doi:10.1111/nph.13973

58. Healey, A., Furtado, A., Cooper, T. & Henry, R. J. Protocol: A simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods* **10**, 1–8 (2014).

59. Luo, R. *et al.* Erratum to 'SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler' [GigaScience, (2012), 1, 18]. *Gigascience* **4**, 1 (2015).

60. Leggett, R. M., Clavijo, B. J., Clissold, L., Clark, M. D. & Caccamo, M. Next clip: An analysis and read preparation tool for nextera long mate pair libraries. *Bioinformatics* (2014). doi:10.1093/bioinformatics/btt702

61. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

62. Roberts, R. J., Carneiro, M. O. & Schatz, M. C. The advantages of SMRT sequencing. *Genome Biol.* (2013). doi:10.1186/gb-2013-14-6-405

63. Chin, C. S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* (2013). doi:10.1038/nmeth.2474

64. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* (2013). doi:10.1093/bioinformatics/btt086

65. Parra, G., Bradnam, K. & Korf, I. CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* (2007). doi:10.1093/bioinformatics/btm071

Acknowledgements

This work was supported by the ANR grant EVOLSYM (ANR-17-CE20-0006-01) to P-M.D, by the Bill & Melinda Gates Foundation as Engineering the Nitrogen Symbiosis for Africa (OPP1172165), by the BBSRC as OpenPlant to G.E.D.O (BB/L014130/1), by the 10KP

initiative (BGI-Shenzhen), by the National Science Foundation (DEB1831428) to F.-W.L and by the Swedish Research Council VR to U.L. (2011-5609 and 2014-522) and to D.M.E (2016-05180). G.V.R is additionally supported by a BBSRC Discovery Fellowship (BB/S011005/1). Part of this work was conducted at the LRSV laboratory, which belongs to the TULIP Laboratoire d'Excellence (LABEX) (ANR-10-LABX-41). We are grateful to the genotoul bioinformatics platform Toulouse Midi-Pyrenees for providing computing and storage resources. We thank Fabrice Roux for helping with the collection of *M. polymorpha* accessions, Aisling Cooke for assistance with *M. paleacea* DNA extraction, Peter Szoevenyi for advices on *M. polymorpha* DNA extraction, David Barker and members of the ENSA project for helpful comments and discussion. Figure 6b was prepared by Jeremy Calli (www.jeremy-calli.fr).

Author contributions

P-M.D., G.V.R., M.K.R., J.K., G.E.D.O. and T.V. conceived the experiments; J.K, H.S.C and L.C. developed symDB; G.V.R., M.K.R., J.K., T.V., D.L.M.M., N.V., C.L., J.C., and P-M.D. conducted the experiments; A-M.L., D.M.E., U.L. generated the *M. polymorpha* subspecies genomes; F-W.L., S.C. and G.K.S.W. generated the *Blasia pusilla* transcriptome; P-M.D., G.V.R., M.K.R., J.K. and T.V. analyzed the data; J.K. compiled the Supplementary material; G.V.R., M.K.R., G.E.D.O. and P-M.D. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper

Figure legends

Figure 1. Conservation of the symbiotic genes in land plants. The tree on the left depicts the theoretical plant phylogeny. The heatmap indicates the phylogenetic pattern for each of the 34 investigated genes. The type of symbiosis formed by each investigated species is indicated by black boxes. AMS: Arbuscular Mycorrhiza Symbiosis; RNS: Root Nodule Symbiosis; OM: Orchid Mycorrhiza; Ericoid-like; Cyanobacteria: association with cyanobacteria; EcM: EctoMycorrhizae.

Figure 2. Maximum-likelihood trees of genes specific to the arbuscular mycorrhizal (a,b,c) or intracellular (d,e,f) symbioses in land plants. a, *STR* (model: TVMe+R5); **b**, *STR2* (SYM+R6); **c**, *RAD1* (TVMe+R5). **d**, *SymRK* (GTR+F+R5); **e**, *CCaMK* (SYM+R6); **f**, *CYCLOPS* (GTR+F+R5). *STR* and *STR2* trees were rooted using their closest paralogs *STR2*

and *STR* respectively. The *RAD1* tree was rooted on the bryophyte clade. *SymRK*, *CCaMK* and *CYCLOPS* trees were rooted on the bryophyte clade. Species name were coloured as follow, black: species with intracellular symbiosis; light red: species without intracellular infection; light grey: species with undetermined symbiotic status. Cyan dots indicate species forming the arbuscular mycorrhizal symbiosis.

Figure 3. Loss of symbiotic genes following mutualism abandonment in Marchantia. a, Ink stained transversal sections of *Marchantia paleacea* (*Marpal*) and *Marchantia polymorpha* subspecies *ruderalis* (*Marpol rud*), *montivagans* (*Marpol mon*) and *polymorpha* (*Marpol pol*). Arbuscules are present in the midrib of *Marchantia paleacea* and absent from the three *Marchantia polymorpha* subspecies. top images, bar= 1 mm; bottom images, bar= 0.25 mm. **b**, *M. paleacea* gene models aligned with the corresponding pseudogenized loci from the three *M. polymorpha* subspecies. **c,** Multiple sequence alignment diversity in a ~1kb region of *CCaMK*. Pseudogenization pattern in 35 *Marchantia polymorpha* accessions compared to the three *Marchantia polymorpha* subspecies. Red vertical lines indicate mismatches and white boxes/red horizontal lines indicate gaps.

Figure 4. Conservation of CCaMK and CYCLOPS biochemical properties in land plants. **a,** *Medicago truncatula pENOD11:GUS* roots transformed with *pUb:CCaMK-K* from *Mimosa pudica* (*Mimpud*), *Discaria trinervis* (*Distri*), *Fragaria vesca* (*Fraves*), *Hordeum vulgare* (*Horvul*), *Zea mays* (*Zeamay*) and *M. paleacea* (*Marpal*) show strong activation of the *ENOD11:GUS* reporter (in blue). Control roots transformed with an empty vector show little or no GUS activity. Numbers of plants showing a strong *ENOD11:GUS* activation out of the total transformed plants are indicated. **b,** *M. truncatula ccamk* mutant roots transformed with *pUb:CCaMK* from *M. pudica*, *D. trinervis*, *F. vesca*, *H. vulgare*, *Z. mays* and *M. paleacea* show infected nodules 26 days post inoculation with *Sinorhizobium meliloti LacZ*. Bacteria in the nodules are stained in blue. A representative infected nodule is shown for each *CCaMK* ortholog. Number of plants showing infected nodules out of the total transformed plants are indicated. Scale bar 200µm. **c,** *M. truncatula cyclops* mutant roots transformed with *pUb:CYCLOPS* from *M. truncatula*, *M. paleacea* and an empty vector (control) show nodules with variable infection level. Whereas with the control plants most of the nodules are unifected or with arrested infection (as illustrated), with MedtruCYCLOPS and MarpalCYCLOPS, fully infected nodules are observed (as illustrated). The boxplot shows differences in the percentage of fully infected nodules per plant ($n_{\_control}$=19, $n_{\_Marpal}$=29, $n_{\_Medtru}$=21). "+" indicates mean value. Different letters indicate different statistical groups after a FDR correction at a 0.95 threshold (Kruskal-Wallis rank sum test; $\chi^2_{\mathrm{Control}-Marpal} = 7.9343$**, $\chi^2_{\mathrm{Control}-Medtru} =$

11.976^{**}, $\chi^2_{Marpal-Medtru} = 2.2817$). The barplot shows percentage of plants with fully infected nodules. Different letters indicate different statistical groups (Chi-Square test of independence; $\chi^2_{\text{Control}-Marpal} = 4.2338^{*}$, $\chi^2_{\text{Control}-Medtru} = 6.5317^{**}$, $\chi^2_{Marpal-Medtru} = 0.25144$).

Figure 5. Maximum-likelihood trees of infection-related genes. **a**, *LIN* and its paralog *LIN-like* (model: GTR+F+R7); **b**, *VAPYRIN* (GTR+F+R6); **c**, *SYN* (TIM3+F+R5). Due to the high duplication of each families, only the angiosperms clade is displayed for *VAPYRIN* and *SYN*; whereas Gymnosperms were conserved for *LIN* and *LIN*-like due to their divergence following the seed plants whole genome duplication event. Full trees are available as Supplementary Figures 21, 22 and 28 . *LIN/LIN-like* tree was rooted on non-seed plants; whereas *VAPYRIN* and *SYN* trees were rooted on *Amborella trichopoda*. Species names were coloured as follow, black: species with intracellular symbiosis; light red: species without intracellular infection; light grey: species with undetermined symbiotic status. Cyan dots outside indicate species forming AMS.

Figure 6. Model for the conservation of symbiotic genes across symbiosis types. a, The Common Symbiosis Pathway genes *SymRK*, *CCaMK* and *CYCLOPS* (CSP) in all land plants. *RAD1*, *STR* and *STR2* are exclusively conserved in species forming the Arbuscular mycorrhizal symbiosis (AMS). The infection-related genes (*i.e. VAPYRIN*, *SYN* or *LIN/LIN-like*) in angiosperms are specific to species forming intracellular symbiosis (black background). Mutualism abandonment (NM, white) or loss of intracellular symbiosis (grey) result in the loss of all these genes. **b,** Schematic representation of transition among symbiotic types and the conservation of the corresponding genes across land plants. Cyano: cyanobacteria association; EcM: EctoMycorrhizae; OM: Orchid Mycorrhiza; RNS: Root-Nodule Symbiosis; ErM: Ericoid Mycorrhiza; ErM-like: Ericoid-like Mycorrhiza.

666

Table 1: Genome assembly statistics for Marchantia species sequenced as part of this study and comparison to the *Marchantia polymorpha* ssp. *ruderalis* TAK1 reference genome

		***Marchantia polymorpha ssp. ruderalis* TAK-1**	***Marchantia paleacea***	***Marchantia polymorpha* ssp. *polymorpha***	***Marchantia polymorpha* ssp. *montivagans***
Assembly size (Mb)		210.6	238.61	222.7	225.7
Scaffolds		2957	22669	2741	2710
N50 length (Kb)		1313.57	77.78	368.25	589.42
BUSCO score	**Complete**	821	817	855	855
	Single Copy	793	790	832	829
	Duplicated	28	27	23	26
	Fragmented	48	53	38	42
	Missing	571	570	547	543
G+C (%)		41.1	40.3	42.2	42.1
Reference		Bowman *et al.*[26]	This study	This study	This study

667

668

669