# 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.



## ▶ To cite this version:

Guru Radhakrishnan, Jean Keller, Mélanie K Rich, Tatiana Vernié, Duchesse Lacour Mbadinga Mbadinga, et al.. An ancestral signalling pathway is conserved in intracellular symbioses-forming plant lineages. Nature Plants, 2020, 6 (3), pp.280-289. 10.1038/s41477-020-0613-7 . hal-02550894

## HAL Id: hal-02550894

## https://hal.inrae.fr/hal-02550894

Submitted on 26 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 # An ancestral signalling pathway is conserved in plant lineages forming intracellular symbioses
2

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

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

8 [2] LRSV, Université de Toulouse, CNRS, UPS, 31326 Castanet-Tolosan, France

9 [3] LIPM, Université de Toulouse, INRA, CNRS, 31326 Castanet-Tolosan, France.

10 [4] Plant Ecology and Evolution, Department of Ecology and Genetics, Evolutionary Biology
11 Centre, Uppsala University, 75236 Uppsala, Sweden

12 [5] Agricultural Genome Institute at Shenzhen, Chinese Academy of Agricultural Sciences,
13 Shenzhen 518124, China

14 [6] BGI-Shenzhen, Shenzhen 518083, China.

15 [7] Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada.

16 [8] Department of Medicine, University of Alberta, Edmonton, AB T6G 2E, Canada.

17 [9] Boyce Thompson Institute, Ithaca, New York, USA.

18 [10] Plant Biology Section, Cornell University, New York, USA.

19 [11] Sainsbury Laboratory, University of Cambridge, Bateman Street, Cambridge CB2 1LR, UK

20 * These authors contributed equally.

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

22 **ABSTRACT**

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

36

37 **Introductory paragraph**

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

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

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

80 Targeted phylogenetic analyses have identified multiple symbiotic genes in the
81 transcriptomes of bryophytes, but study on the overall molecular conservation of symbiotic
82 mechanisms in land plants are lacking[22–24]. Similarly, knowledge on the plant molecular
83 mechanisms behind the diverse array of mutualistic associations, either intracellular or
84 extracellular, are poorly understood[10]. Here, we demonstrate through analysis of a
85 comprehensive set of plant genomes and transcriptomes, the loss and conservation of
86 symbiotic genes associated with the evolution of diverse mutualistic symbioses in plants.
87

88 ## RESULTS AND DISCUSSION

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

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

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

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

145 mutations, deletions and insertions (Fig. 3) and their presence supports a recent abandonment
146 of mutualism in *M. polymorpha*. To better position the timing of this abandonment, we collected
147 35 *M. polymorpha* ssp. *ruderalis* accessions in Europe (Supplementary Table 4) and
148 sequenced *CYCLOPS* and *CCaMK*. All accessions harboured pseudogenes at these two loci,
149 confirming the fixation of these null alleles in the subspecies *ruderalis* (Fig. 3 and
150 Supplementary Figure 36). Besides *M. polymorpha* ssp. *ruderalis*, two other *M. polymorpha*
151 subspecies have been reported, ssp. *polymorpha* and ssp. *montivagans,* that are sister to *M.*
152 *paleacea*. We phenotyped these two subspecies in controlled conditions, and confirmed that
153 the loss of the AM symbiosis occurred before the radiation of the three *M. polymorpha*
154 subspecies approximately 5 million years ago[28] (Fig. 3). We sequenced high-quality genomes
155 of *M. polymorpha* ssp. *montivagans* and *polymorpha* and searched for the presence of the six
156 aforementioned genes. As for *M. polymorpha* ssp. *ruderalis* all six genes were pseudogenized
157 or missing in these two novel assemblies (Fig. 3). As expected for genes under relaxed
158 selection, the signatures of pseudogenization were different between the three subspecies
159 (Fig. 3).

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

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

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

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

204 All other lineages that switched from AMS to other types of mutualistic symbioses
205 retained the three signalling genes *SymRK, CCaMK and CYCLOPS* (Fig. 2, Supplementary
206 Fig. 13–15). These lineages are scattered throughout the land plant phylogeny, thus excluding
207 the hypothesis of a lineage-specific retention of these genes. The three genes were found in
208 the genomes of three Orchidaceae, *Apostasia shenzhenica*, *Dendrobium catenatum* and
209 *Phalaenopsis equestris* and in the transcriptome of *Bletilla striata*[37], that form Orchid
210 Mycorrhizae with Basidiomycetes that develop intracellular pelotons[5]. All three genes were
211 also detected in the transcriptomes of liverworts from the Jungermaniales order *Scapania*
212 *nemorosa* (3/3), *Calypogeia fissa* (1/3), *Odontoschisma prostratum* (1/3), *Bazzania trilobata*
213 (2/3) and *Schistochila sp.* (1/3), which have switched from AMS to diverse Ericoid-like
214 associations with Basidiomycetes and Ascomycetes that form intracellular coils[5]. Furthermore,
215 these genes are also conserved in the legume genus *Lupinus,* which associates with nitrogen-
216 fixing rhizobia forming intracellular symbiosomes in root nodules, but have lost AMS[38].

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

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

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

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

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

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

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

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

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

308 **CONCLUDING REMARKS**

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

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

326 **Methods**

327 **Genome and transcriptome sequencing**

328 ***Marchantia paleacea***

329 Plant material. *M. paleacea* thalli previously collected in Mexico (Humphreys) were kindly
330 provided by Dr Katie J. Field and Professor David J. Beerling (The University of Sheffield).
331 Gemma from these thalli were collected using a micropipette tip and placed into a micro-
332 centrifuge tube. 5% sodium hypochlorite solution was used to sterilize the gemmae for about
333 30 seconds followed by rinsing with sterile water 5 times to remove residual sodium
334 hypochlorite solution. The sterilized gemmae were grown on Gamborg's half-strength B5
335 medium on sterile tissue culture plates under a 16/8-hour day-night cycle at 22° C under
336 fluorescent illumination with a light intensity of 100 μmol/μm$^2$s.

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

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

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

359 *Genome assembly.* Adapter and quality trimming were performed on the paired-end library
360 using Trimmomatic v0.33 using the following parameters (ILLUMINACLIP:TruSeq2-
361 PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:12). The trimmed
362 paired-end reads were carried forward for assembling the contigs using multiple assemblers.
363 Scaffolding of the contigs was done using the scaffolder of SOAPdenovo2[59] with the mate-pair
364 library after processing the reads through the NextClip[60] pipeline to only retain predicted
365 genuine long insert mate-pairs. Assembly completeness was measured using the BUSCO[61]
366 plants dataset.
367
368 *Transcriptome sequencing.* cDNA was purified and concentrated on MinElute Columns
369 (Qiagen) and used to construct an Illumina library using the Ovation Rapid DR Multiplex
370 System 1-96 (NuGEN, Redwood City, CA). The library was amplified using MyTaq (Bioline,
371 London, UK) and standard Illumina TruSeq amplification primers. PCR primer and small
372 fragments were removed by Agencourt XP bead purification. The PCR components were
373 removed using an additional purification on Qiagen MinElute Columns. Normalisation was
374 done using Trimmer Kit (Evrogen, Moscow, Russia). The normalized library was re-amplified
375 using MyTaq (Bioline, London, UK) and standard Illumina TruSeq amplification primers. The
376 normalized library was finally size selected on a LMP-Agarose gel, removing fragments smaller
377 than 350Bp and those larger than 600Bp. Sequencing was done on an Illumina MiSeq on 2 x
378 300bp mode. The RNA extractions, cDNA synthesis library preparation and sequencing were
379 carried out by LGC Genomics GmbH (Berlin, Germany).
380
381 ***Blasia pusilla***
382 *Blasia pusilla* was originally collected from Windham County, Connecticut, USA, and
383 maintained in Duke University greenhouse. RNA was extracted from plants with symbiotic
384 cyanobacterial colonies using Sigma Spectrum Plant Total RNA kit. Library preparation and
385 sequencing were done by BGI-Shenzhen. A Ribo-Zero rRNA Removal Kit was used to prepare
386 the transcriptome library. In total, 3 libraries were constructed, which were sequenced on the
387 Illumina Platform Hiseq2000, 150bp paired-ends, with insert size 200bp.
388
389 ***Marchantia polymorpha subspecies***

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

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

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

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

399 Supplementary material and method information is available for this paper.

400 **Data availability**

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

406 **References**


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

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

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

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

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

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

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

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

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

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

428 11. van Velzen, R. *et al.* Comparative genomics of the nonlegume Parasponia reveals
429 insights into evolution of nitrogen-fixing rhizobium symbioses. *Proc. Natl. Acad. Sci.*
430 201721395 (2018). doi:10.1073/pnas.1721395115

431 12. Albalat, R. & Cañestro, C. Evolution by gene loss. *Nat. Rev. Genet.* **17**, 379–391
432 (2016).

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

435 14. Delaux, P. M. Comparative phylogenomics of symbiotic associations. *New Phytol.*
436 **213**, 89–94 (2017).

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

440 16. Bravo, A., York, T., Pumplin, N., Mueller, L. A. & Harrison, M. J. Genes conserved for
441 arbuscular mycorrhizal symbiosis identified through phylogenomics. *Nat. Plants* **2**,
442 (2016).

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

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

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

450 20. Keymer, A. *et al.* Lipid transfer from plants to arbuscular mycorrhiza fungi. *Elife* **6**,
451 (2017).

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

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

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

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

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

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

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

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

473 29. Read, D. J. & Perez-Moreno, J. Mycorrhizas and nutrient cycling in ecosystems - A
474 journey towards relevance? *New Phytologist* (2003). doi:10.1046/j.1469-
475 8137.2003.00704.x

476 30. Rey, T. *et al.* The Medicago truncatula GRAS protein RAD1 supports arbuscular
477 mycorrhiza symbiosis and *Phytophthora palmivora* susceptibility. *J. Exp. Bot.* **68**,
478 5871–5881 (2017).

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

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

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

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

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

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

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

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

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

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

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

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

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

509 43. Buendia, L., Wang, T., Girardin, A. & Lefebvre, B. The LysM receptor-like kinase
510 SlLYK10 regulates the arbuscular mycorrhizal symbiosis in tomato. *New Phytol.* **210**,
511 184–195 (2016).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

545 57. Huisman, R. *et al.* A symbiosis-dedicated SYNTAXIN OF PLANTS 13II isoform
546 controls the formation of a stable host-microbe interface in symbiosis. *New Phytol.*
547 (2016). doi:10.1111/nph.13973

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

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

553 60. Leggett, R. M., Clavijo, B. J., Clissold, L., Clark, M. D. & Caccamo, M. Next clip: An
554 analysis and read preparation tool for nextera long mate pair libraries. *Bioinformatics*
555 (2014). doi:10.1093/bioinformatics/btt702

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

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

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

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

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

568

569 **Acknowledgements**

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

573 initiative (BGI-Shenzhen), by the National Science Foundation (DEB1831428) to F.-W.L and
574 by the Swedish Research Council VR to U.L. (2011-5609 and 2014-522) and to D.M.E (2016-
575 05180). G.V.R is additionally supported by a BBSRC Discovery Fellowship (BB/S011005/1).
576 Part of this work was conducted at the LRSV laboratory, which belongs to the TULIP
577 Laboratoire d'Excellence (LABEX) (ANR-10-LABX-41). We are grateful to the genotoul
578 bioinformatics platform Toulouse Midi-Pyrenees for providing computing and storage
579 resources. We thank Fabrice Roux for helping with the collection of *M. polymorpha* accessions,
580 Aisling Cooke for assistance with *M. paleacea* DNA extraction, Peter Szoevenyi for advices on
581 *M. polymorpha* DNA extraction, David Barker and members of the ENSA project for helpful
582 comments and discussion. Figure 6b was prepared by Jeremy Calli (www.jeremy-calli.fr).

583 **Author contributions**

584 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
585 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.
586 conducted the experiments; A-M.L., D.M.E., U.L. generated the *M. polymorpha* subspecies
587 genomes; F-W.L., S.C. and G.K.S.W. generated the *Blasia pusilla* transcriptome; P-M.D.,
588 G.V.R., M.K.R., J.K. and T.V. analyzed the data; J.K. compiled the Supplementary material;
589 G.V.R., M.K.R., G.E.D.O. and P-M.D. wrote the manuscript.

590 **Competing interests**

591 The authors declare no competing interests.

592 **Additional information**

593 Supplementary information is available for this paper

594 **Figure legends**

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

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

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

610

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

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

640 $11.976^{**}$, $\chi^2_{Marpal-Medtru} = 2.2817$). The barplot shows percentage of plants with fully infected
641 nodules. Different letters indicate different statistical groups (Chi-Square test of independence;
642 $\chi^2_{\text{Control}-Marpal} = 4.2338^{*}$, $\chi^2_{\text{Control}-Medtru} = 6.5317^{**}$, $\chi^2_{Marpal-Medtru} = 0.25144$).

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

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

663

664

665

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