# Leafy liverwort genomes shed light on the evolution of ericoid symbiosis

Leonardo Castanedo[1], Matheus E. Bianconi[1], Tifenn Pellen[1], Tatiana Vernié[1], Corinne Lefort[1], Cyril Libourel[1+], Aurélie Le Ru[2], Nathalie Rodde[3], Stephane Cauet[3], Caroline Callot[3], Laura Forrest[4], Tomoaki Nishiyama[5,6], David Wickell D[7], Fay-Wei Li[7,8], Jean Keller[1#], Pierre-Marc Delaux[1#]

[1] LRSV, Universite de Toulouse, CNRS, UPS, Toulouse INP, 31326 Castanet-Tolosan, France.

[2] Plateforme d'Imagerie FRAIB-TRI, Universite de Toulouse, CNRS, Auzeville-Tolosane 31320, France.

[3] INRAE, CNRGV French Plant Genomic Resource Center, F-31320 Castanet-Tolosan, France.

[4] Royal Botanic Gardens Edinburgh, 20A Inverleith Row, Edinburgh, EH3 5LR, UK.

[5] Research Center for Experimental Modeling of Human Disease, Kanazawa University, Kanazawa, Ishikawa, Japan.

[6] School of Science, Academic Assembly, University of Toyama, Toyama, Toyama, Japan

[7] Boyce Thompson Institute, Ithaca, NY, USA

[8] Plant Biology Section, Cornell University, Ithaca, NY, USA

[+]present address: Unité de Recherche Physiologie, Pathologie, et Génétique Végétales (PPGV), INP PURPAN, Université de Toulouse, Toulouse, France

*correspondance: pierre-marc.delaux@cnrs.fr and jean.keller@cnrs.fr

**Abstract**

Mycorrhizal symbiosis has been fundamental to colonization of terrestrial ecosystems by plants, influencing population dynamics, biogeography, and their evolution[1–3]. The emergence of the ancient arbuscular mycorrhizal symbiosis (AMS) was followed by the diversification of alternative mycorrhizal types, including the ericoid mycorrhizae (ErM)[3]. The evolution of ErM coincides with the emergence of the Ericaceae family of angiosperms and the diversification of leafy liverworts within the order Jungermanniales, as well as their expansion into boreal and high-altitude ecosystem[4–6]. Despite their evolutionary and ecological relevance, little is known about the molecular and genomic basis of ErM. Here, we sequenced the nuclear genomes of two leafy liverworts from Jungermanniales, and reconstituted ErM under laboratory conditions. We demonstrated the existence of a nutrient-regulated symbiotic state that enables ErM and underlies intracellular colonization of plant tissues. Genome and functional analyses revealed that leafy liverworts retain a conserved

suite of functional genes that are required for endosymbiosis in land plants. Comparative transcriptomic analysis under ErM-permissive and restrictive conditions identified gene modules associated with endosymbiosis. Transcriptome sequencing cross-comparisons of ErM and AMS identified a conserved gene module required for the intracellular accommodation of fungal symbionts, marking it as a shared feature across diverse types of land plant symbioses. Our study presents the first genomic resources for leafy liverworts, and provides the first experimental evidence that ErM evolved from AMS. This evolutionary transition involved the co-option of a nutrient-based mechanism for the establishment of ErM, and additionally the co-option of two gene modules essential for signaling and symbiont accommodation during endosymbiosis. Our data offer valuable insights and resources for understanding the molecular basis of ErM and its diversification from AMS in land plants.

**Main**

Symbiotic interactions, particularly mycorrhizal symbiosis (*i.e.,* mycorrhizae), have played a fundamental role in the colonization of terrestrial ecosystems by plants[1–3]. Therefore, mycorrhizae can influence both population dynamics and biogeographical distribution, thus shaping plant evolution[3]. The appearance of the first mycorrhizal type, the arbuscular mycorrhiza symbiosis (AMS), coincides with the earliest land plant macrofossil records dating back to 400 million years ago (Ma)[7–9]. This ancient origin has been corroborated by evo-devo and phylogenetic studies[10–12]. The evolution of AMS in land plants was followed by the diversification of alternative mycorrhizal types and the origin of novel plant clades that colonized previously unoccupied terrestrial ecosystems[3]. For example, the evolution of ericoid mycorrhizae (ErM) around 140–75 Ma coincided with the origin and the appearance of the Ericaceae family of vascular plants in boreal and high-altitude forests, as well as in heathland biomes[3–5]. ErM can also be found in bryophytes, within the leafy liverworts order Jungermanniales[13,14]. Liverworts are one of the three clades of bryophytes and comprise *c.a.,* 5,000 species[15] classified into three main classes: Haplomitriopsida, Marchantiopsida (also called complex thalloid) and Jungermanniopsida, containing both simple thalloid and leafy liverworts, including the ErM order Jungermanniales. Although Jungermanniopsida is the most diverse class of liverworts, encompassing over 4,000 species, it remains underrepresented in genomic datasets with no whole genome sequences currently available. This contrasts with the complex thalloid class, in which species such as *Marchantia polymorpha* have been extensively studied[16–19]. It is estimated that ErM in Jungermanniales, alongside Ericaceae plants, contribute to the mobilization of nitrogen (N) and phosphate (P) in the habitats where these plants often co-exist[20,21,6]. Even though ErM appears to have

diversified from AMS in liverworts[11], the cellular, molecular and genetic mechanisms underlying this diversification event remain unclear.

In this study, we sequenced the genome of two leafy liverworts from the order Jungermanniales (*i.e., Calypogeia fissa* and *Solenostoma infuscum*). By conducting inoculation experiments with the known ErM-forming fungus *Hyaloscypha hepaticicola* (syn. *Rhizoscyphus ericae*), we reconstituted the ErM in laboratory conditions and demonstrated the occurrence of nutrient-regulated permissive symbiotic states that govern the establishment of endosymbiosis. The genome of leafy liverworts contains the suite of genes with well-known and established functions during plant endosymbioses. Cross-comparison of transcriptomic data in *C. fissa* inoculated with *H. hepaticicola* under ErM-permissive and restrictive conditions, identified genetic modules associated with early signaling events during endosymbiosis. Finally, cross-comparisons of transcriptomic data between ErM- and AMS-permissive conditions identified a gene module that includes a LysM-RLK receptor, a transcription factor, and two infectosome-related genes as the hallmark of endosymbiosis in land plants.

## Results and Discussion

### Sequencing of leafy liverwort genomes

In order to provide genomic resources for Jungermanniopsida to further study ErM, we produced genome assemblies for two species, *Solenostoma infuscum* (previously referred to as *Jungermannia infusca*) and *Calypogeia fissa.* The *S. infuscum* genome was sequenced using the Illumina technology, yielding 248 million of paired-end reads assembled into 4,616 scaffolds (from 949bp to 1,4Mb, N50=244Kb) (Table 1). The *C. fissa* genome was sequenced using the PacBio long-read technology, resulting in an assembly comprising 82 scaffolds (from 46Kb to 66Mb, N50=16Mb) (Table 1). Whereas the estimated size of the *S. infuscum* genome is *ca.* 400 Mb, the estimated genome size of *C. fissa* is *ca.* 800 Mb. Analysis assessing the completeness of the assembled genomes using the Viridiplantae odb10 BUSCO database indicated 95.3% and 96% completion for *S. infuscum* and *C. fissa*, respectively, with 93.9% of the annotated genes duplicated in *C. fissa* (Table 1). Gene structural annotation predicted 14,494 and 38,150 protein-coding genes in the genomes of *S. infuscum* and *C. fissa* respectively (Table 1). The number of predicted genes in the *C. fissa* genome is approximately twice the average number of protein-coding genes found in other liverworts (*ca.* 19,000). Overall, the genome assembly statistics supported the hypothesis that *C. fissa* underwent whole-genome duplication (WGD), as previously proposed[15]. This hypothesis is further supported by the distribution of synonymous substitutions per site (Ks) between pairs of paralogs in *C. fissa,* with a single peak near the

origin (Ks = 0.05), which is indicative of a recent large-scale duplication event (Fig. 1A). Likewise, an intragenomic synteny analysis recovered 137 collinear blocks comprising 7,087 gene pairs (*i.e.,* syntelogs) representing 37.2% of the *C. fissa* genome (Fig. 1B), with nearly all of these corresponding to gene pairs (Fig. 1A, Ks = 0.05). In addition to paralogous Ks, orthologous Ks was estimated for *C. fissa* and *S. infuscum*. The resulting Ks plot of orthologous genes exhibited a peak around Ks = 0.35, which is largely consistent with a scenario of recent WGD in *C. fissa*, long after the divergence between the two species (Supplementary Fig. 1). Intergenomic comparison of synteny between *C. fissa* and the distantly related species *M. polymorpha* identified relatively few collinear gene pairs between them, roughly 10% of genes in each species, likely owing to the species' deep phylogenetic divergence (Supplementary Fig. 2). However, a clear 2:1 pattern of syntenic depth was observed among collinear genes (Fig. 1C) consistent with the apparent absence of WGD events in *M. polymorpha*.

The species tree inferred from the orthologous groups recovered for the selected bryophyte species in the data set was congruent to the currently accepted topology for the group[22], with *C. fissa* and *S. infuscum* forming a clade that is sister to the Marchantiopsida clade of complex thalloid liverworts (Fig. 1D). Using molecular dating analysis with secondary calibration based on published data[23], we estimated the divergence between *C. fissa* and *S. infuscum* at 69.8 Ma (95% HPD = 32 - 122 Ma), and that between Jungermaniopsida and Marchantiopsida at 293.7 Ma (95% HPD = 217 - 370 Ma; Fig. 1D).

Here, we have generated two genome assemblies for the Jungermanniales, in which further analysis indicates a recent whole genome duplication in *C. fissa*.

### ***Calypogeia fissa* establishes a mutualistic intracellular symbiosis with *Hyaloscypha hepaticicola***

With the genomes of leafy liverworts now available, we aimed at developing a model for studying ErM in controlled conditions. To do this, we conducted experiments in which we cultured mock-inoculated, or inoculated *C. fissa* plants with the ErM-forming fungus *Hyaloscypha hepaticicola*, in a nutrient regime low in both phosphate (P) and nitrogen (N) (Fig. 2). We used this treatment to mimic the nutrient-limited conditions in which leafy liverworts can thrive, and to test the hypothesis that ErM contributes to plant resilience in environments with low P and N[20]. Two parameters indicative of mutualism were scored: i) plant growth and ii) rhizoid-tip colonization (Fig. 2A-P). By comparison to mock-inoculated treatments in which no fungal growth was detected, the growth of *H. hepaticicola* in *C. fissa* plants was discrete and highly localized around and inside swollen rhizoid tips of *C. fissa* (Fig. 2G-H, and 2L-2M). We observed that inside the swollen rhizoid tips of *C. fissa*

intracellular coils of *H. hepaticicola*-hyphae were developed (Fig. 2N-O). The development of hyphal coils upon ErM are considered the symbiotic structures that mediate the exchange of nutrients between the plant host and ericoid mycorrhizae-forming fungus in Jungermanniales[13,24]. To assess whether colonization benefited *C. fissa* plants, we quantified plant growth in both *H. hepaticicola*-inoculated and mock-inoculated conditions over a 9-week time-course experiment (Fig. 2A-B, and 2E). We observed that inoculated *C. fissa* plants qualitatively and quantitatively produced substantially more biomass compared to mock-inoculated plants (Fig. 2A-B, and 2E).

Next, the interaction between *C. fissa* and *H. hepaticicola* was assessed in extreme nutrient depletion conditions (Fig. 2C-D). For this, growth of mock-inoculated or inoculated *C. fissa* plants was determined over a 9-week time-course experiment in a condition supplied only with water (Fig. 2F). We observed that inoculated *C. fissa* plants produced more biomass compared to non-inoculated plants, although the difference was not statistically significant (Fig. 2C-D, and 2F). In contrast to inoculated plants which appeared greener, mock-inoculated plants displayed symptoms of nutrient deficiency stress, such as stem chlorosis and senescence (Fig. 2C-D). This suggests that the fungus *H. hepaticicola* might provide stress resilience to *C. fissa* under extreme nutrient deficiency conditions. Similar to our observations under low P/N supply, inoculated *C. fissa* plants developed a high number of swollen rhizoid tips per stem that were intracellularly colonized by *H. hepaticicola* (Fig. 2P).

In plate experiments, we observed that *H. hepaticicola* produced a small ring of mycelium when grown under low P/N or water-only conditions (Supplementary Fig. 3). This contrasts with the localized, yet robust fungal growth inside and around the rhizoids of *C. fissa* in both low P/N, and in water-only conditions (compare Fig. 2G-O with Supplementary Fig. 3F-G). These results suggest that decreased growth of *H. hepaticicola* in limiting nutrient medium on plates is compensated when the fungus is grown in the presence of *C. fissa* indicative of a mutually beneficial interaction. However, the benefit obtained by *H. hepaticicola* requires further experimental testing.

These results indicate that our experimental system recapitulates ErM in *C. fissa* which is marked by intracellular colonization of rhizoid tips. By comparison to mock-inoculated plants, intracellular colonization in *C. fissa* coincides with a substantial growth benefit under extremely low nutrient regimes.

**A symbiotic permissive state defined by nutrient availability is a conserved trait across mycorrhizal symbioses**

Given that P and N availability regulates the establishment of AM in angiosperms and thalloid liverworts, promoting a symbiotic permissive state[12,25,26], and based on the observations made in this study, it can be hypothesized that both P and N availability might regulate the establishment of ErM in leafy liverworts such as *C. fissa*.

To test whether P supply alone, as in AM-forming species, governs the establishment of ErM in *C. fissa*, we conducted inoculation experiments under a range of P supply (Fig. 3, Supplementary Figs. 4 and 5). Before this, we determined whether any of the tested nutrient conditions had a detrimental effect on the growth of *H. hepaticicola* (see Supplementary Fig. 3, and Supplementary Table 1). Although fungal growth varied according to the nutrient regime, none of the tested conditions abolished it (Supplementary Fig. 3). We observed that *H. hepaticicola* produced a well-defined ring of melanized and compacted mycelium under high N conditions which was unaffected by P supply (Supplementary Fig. 3A, 3C, and 3H). The growth phenotype of *H. hepaticicola* under high N supply has been observed on commercial medium and is typically displayed by Helotialean fungi growing on plates with replete nutrient conditions[27–30].

We next applied these nutrient conditions to mock- or *H. hepaticicola*-inoculated *C. fissa* plants. Under conditions of low P supply, *C. fissa* inoculated with *H. hepaticicola* produced more biomass than mock-inoculated plants (Fig. 3D-F, and Supplementary Fig. 4A-C, see Supplementary Table 2). This is in contrast with the minimal, or none, growth differences observed between inoculated and mock-inoculated *C. fissa* plants under high P supply (Fig. 3A-F, Supplementary Fig. 5A-F). Whereas inoculated *C. fissa* plants only show a modest increase in biomass production under conditions of high P and high N supply, this effect was suppressed by reducing N supply (Supplementary Fig. 5A-F). Consistent with these results, we determined positive mycorrhizal growth responses of *C. fissa* plants under P deficiency, contrasting with the minimal or negative mycorrhizal growth responses under high P and limited N supply (see Supplementary Table 2).

Under low P supply we observed the development of swollen rhizoid tips that contained intracellular hyphae (Fig. 3G-M, Supplementary Fig. 4F-J, Suppl. Table 3). Contrasting these observations, significantly fewer swollen rhizoid tips, or none at all, were observed under conditions of high P and limited N supply (Fig. 3M).

The nutrient-mediated permissive state for the establishment of AM in angiosperms and thalloid liverworts is mediated by transcriptional regulation under the control of upstream factors of the PHOSPHATE STARVATION RESPONSE, such as PHR1, which are conserved in all land plants, including *C. fissa*[31]. In addition to AM, the nitrogen-fixing root nodule endosymbiosis formed by few angiosperm species and nitrogen-fixing bacteria, is

regulated by N availability[32]. Here we provide experimental evidence indicating that a third example of nutrient-regulated endosymbiosis is exemplified by ErM which is established upon P limitation and fine-tuned by N supply. We propose that the regulation of colonization by the plant nutrient status is a key driver in maintaining mutualism and may be just as important as the reciprocal reward itself[33].

Our data indicates that a symbiotic permissive state for the establishment of ErM is directly regulated by the supply of both P and N in *C.fissa*.

**Leafy liverwort genomes harbor functional endosymbiosis-related genes**

The establishment of the AM relies on a set of three genes (SYMRK, CCaMK and CYCLOPS), which constitute the Common Symbiosis Pathway (CSP)[34,35]. *SYMRK* encodes a receptor-like kinase, and *CCaMK* encodes a kinase that mediates the phosphorylation of the transcription factor CYCLOPS, which in turn activates the expression of AM-related genes in land plants[34]. The CSP genes are both conserved and required for the establishment of the AM across land plants, including angiosperms and bryophytes[10,35,36]. Consistent with this, plant lineages that have lost the ability to form AM also lost the CSP during evolution[10,37]. Very few exceptions to these correlated losses have been observed, and they always correspond to the occurrence of another type of intracellular symbiosis, such as the nitrogen-fixing root nodule symbiosis in *Lupinus*[11]. Preliminary observations in the transcriptomes of Jungermanniales liverworts suggested the presence of transcribed CSP genes[11]. By contrast, other genes such as the ones encoding the half-ABC transporters STR and STR2, which are presumably involved in the transfer of lipids from host plants to the AM fungi, were not detected among the transcriptomes of land plants other than AM-forming plant linages[11]. Here we searched for the presence of the CSP and 13 other genes associated with AM in the genomes of *C. fissa* and *S. infuscum* (Fig. 4A, Supplementary Fig. 6).

By contrast with our expectations, in the genome of both *C. fissa* and *S. infuscum* we detected the presence of three genes with high sequence homology to genes required for the establishment of AM, such as the GRAS transcription factor *RAD1*, and the two half-ABC transporters *STR* and *STR2* (Supplementary Fig. 6). The apparent absence of these three putative ortholog genes in the previously analyzed transcriptomes of leafy liverworts[11] suggests that they might be expressed at very low level in these species. An analysis of potential shifts in selection on RAD1, STR and STR2 using the ratio of nonsynonymous to synonymous substitution rates suggests relaxation of selection in *STR* and *STR2*, but not in RAD1 (Supplementary Table 4). Based on this observation, it can be postulated that relaxed selection on these two genes, resulting from the loss of AM, might represent a step towards

gene pseudogenization or neofunctionalization, as observed for some genes of plant species that have lost the capacity to establish AM[11].

We detected orthologs for all three genes (*SYMRK, CCaMK* and CYCLOPS) of the CSP in the genomes of both *C. fissa* and *S. infuscum* (Fig. 4A). In diverse AM-forming species, the CYCLOPS protein has been found to activate gene expression via specific binding to specific *cis*-elements following its phosphorylation by CCaMK[35,38,39]. To determine whether phosphorylation of CYCLOPS by CCaMK during symbiosis is conserved in Jungermanniales, we conducted transactivation assays with the *CYCLOPS* ortholog genes of *C. fissa* (*CfCYCLOPSa* and *CfCYCLOPSb*) and another leafy liverwort, *Scapania nemorea* (*SnCYCLOPS*). As a control, we used *CYCLOPS* from *Medicago truncatula*. We expressed either *CfCYCLOPS* (each of the two paralogs Cf*CYCLOPSa* and Cf*CYCLOPSb*) or *SnCYCLOPS* in combination with an autoactive version of CCaMK (CCaMK-Kin)[40] and the reporter construct *CYC-RE$_{pro}$:GUS*[35,39] in *N. benthamiana* leaves. If functional, the autoactive CCaMK-Kin protein is expected to phosphorylate CYCLOPS leading to transactivation of the co-transformed *CYC-RE$_{pro}$:GUS* reporter construct. As expected, no GUS activity was detected in *N. benthamiana* leaves transformed with *CYC-RE$_{pro}$:GUS* with and without *MtCCaMK-Kin* (Fig. 4B). Contrasting this, strong GUS activity was observed upon co-transformation of the *MtCCaMK-Kin* construct with either *CfCYCLOPS* (a and b), *SnCYCLOPS*, or with the positive *MtCYCLOPS* control construct.

The retention of the endosymbiotic CSP gene module in *C. fissa* and *S. infuscum*, despite the loss of AM in both species, and our functional data in *C. fissa*, support the role of the CSP module for the establishment of ErM.

**Ericoid mycorrhiza leads to a specific transcriptional reprogramming**

In land plants, association with symbiotic organisms lead to the transcriptional induction of gene modules involved at different steps during endosymbiosis[12,41–43]. To gain insight into the molecular mechanisms governing ErM, we conducted bulk RNA sequencing on mock treated or *H. hepaticicola*-inoculated *C. fissa* plants. For this, plants were grown in a permissive condition and harvested on week 8-post inoculation (Supplementary Table 1). Although our data indicated that maximum permissive levels for symbiosis were achieved in a condition with low P/N supply (see Supplementary Table 2), the resulting plant RNA yield was insufficient for downstream sequencing. Thus, we selected the second most permissive condition for ErM with a low and a moderate supply of P and N respectively (Supplementary Tables 1 and 2). Then, differentially expressed genes were computed and identified. Among the differentially regulated genes, we observed 787/743 genes up-/down-regulated in

inoculated plants compared to the mock condition in this permissive condition for ErM (Figure 5A, Supplementary Dataset 1).

Having defined nutrient conditions governing permissive or non-permissive symbiotic states, we hypothesized that comparing the transcriptomic response of *C. fissa* to the inoculation with *H. hepaticicola* in each of these contrasting states could identify endosymbiosis-specific genes modules during ErM (Fig. 5A-B). It is noteworthy to mention that such an experiment cannot be conducted in the context of the AM as the fungal symbionts are obligate biotrophs that display limited growth in the absence of a plant host[44]. Thus, we conducted RNA sequencing experiments with mock-inoculated or *H. hepaticicola*-inoculated *C. fissa* plants that were grown under a restrictive nutrient regime with high P supply regime for 8 weeks (Supplementary Table 1). In this ErM-restrictive condition, where no intracellular colonization was detected (see Fig. 3A-C, 3G-H, and 3M) we computed 566/27 genes up/downregulated in the inoculated plants compared to mock-inoculated plants (Fig. 5B, Supplementary Dataset 2). To identify the genes and processes associated with intracellular symbiosis during ErM, we compared i) the list of DEG in the permissive versus non-permissive conditions, and ii) the functional gene ontology enrichment in each condition (Fig. 5A-B, Supplementary Dataset 3). These comparisons revealed 82/1 genes up/downregulated in response to *H. hepaticicola* that were specifically responsive during the ErM-permissive state.

Among the symbiosis-specific upregulated genes, we observed two highly enriched gene ontology categories. The first category is represented by a large number of enzymes involved in cell-wall remodeling, including eight Xyloglucan Transferases, a Glycosyl Transferase 92, a pectin Lyase and multiple Glycosyl Hydrolases which were specific to the ErM-permissive condition (Fig. 5A, Supplementary Dataset 1). It is noteworthy to mention that the GH81 family of Glycosyl Hydrolase was detected in our analysis. Members of the GH81 family has been described for its function in facilitating infection by endophytic and intracellular symbionts in barley by mitigating cell-wall immune responses[45]. We interpreted these observations to suggest the establishment of ErM in *C. fissa* might involve both the activation of cell-wall modifying enzymes and plant defense modulating responses. The second category includes genes involved in nutrient transport. Among the enriched functional domains are the ‘Glycosyl Hydrolase 32’ and ‘sugar/inositol transporters’. Whilst multiple genes in both categories are upregulated in *C. fissa* under the ErM-permissive condition, none are upregulated in the restrictive condition (Fig. 5A-B, Supplementary Datasets 1 and 2). The glycosil hydrolase 32 correspond to a family of invertase proteins are predicted to break-down sucrose into glucose and fructose. Sugar/inositol transporter proteins are predicted to mediate sugar transport across cell membranes. By contrast, the genes encoding putative lipid transporters STR and STR2, and which were observed to be

under relaxed selection in *C. fissa*, were not detected as transcriptionally responsive under either condition (Supplementary Dataset 1 and 3). These observations suggest that sugar, rather than lipids, is the main carbon source transferred to the ErM-forming fungus *H. hepaticicola* by *C. fissa.*

Another gene category enriched specifically in the ericoid mycorrhiza-permissive condition is 'Ammonium transporter'. Among specific ammonium transporters are the co-orthologs of *AMT2*, an ammonium transporter that has been observed to be both, consistently upregulated during AM in land plants[12], and contributing to ammonium acquisition during endosymbiosis[46]. Interestingly, phosphate transporter proteins from the PHT1 family, which are upregulated during AM, were found to be transcriptionally unresponsive in the ErM-permissive condition in *C. fissa* (Supplementary Dataset 1 and 3), suggesting the involvement of other transporter proteins in nutrient transfers during ErM. A number of genes encoding transporters of the Major Facilitator Superfamily were found upregulated in the symbiotic permissive condition and represent potential candidates.

Our analyses suggest specific transcriptional responses during ErM in *C. fissa*, including mechanisms related to cell wall remodeling and nutrient exchange. These mechanisms are analogous to those in other plant endosymbioses, such as AM. The specific transcriptionally responsive genes detected during ErM suggest a convergent model for the recruitment of processes enabling plant endosymbiosis across different mycorrhizal types.

**Intracellular symbiosis is associated to the universal induction of a specific gene module**

The conservation of several genes required for endosymbiosis across land plants in *C. fissa* and *S. infuscum* suggests that ErM in Jungermanniales evolved from AM. In this context, the evolutionary transition towards ErM might have involved the recruitment of gene modules that enable AM. To further test this hypothesis, we cross-referenced the specific transcriptomic signature of *C. fissa* in response to *H. hepaticicola* under the ErM-permissive condition to the transcriptomic response of the thalloid liverwort *Marchantia paleacea* colonized by the AM-forming fungus *Rhizophagus irregularis*[12]. To conduct this analysis, we used Orthofinder for genome-wide prediction of ortholog genes between both species, following by phylogeny reconstruction of specific genes which were of particular interest due to their functions during AM (see Methods). This analysis identified a total of 20 genes in *C. fissa* upregulated during ErM that had ortholog genes upregulated during AM in *M. paleacea*. Among the genes with known function during endosymbiosis, we detected, *EPR3* (Calfis_g18879 & Calfis_g34255), VAPYRIN (Calfis_g4017 & Calfis_g10154) and *LIN* (Calfis_g31157 & Calfis_g22484) (Supplementary Dataset 4).

The LysM-RLK receptor-encoding *EPR3* gene is required for the recognition of fungal symbionts at the cell surface in legumes[47,48], however, the ligand of EPR3 remains unknown. *VAPYRIN* and *LIN* are genes encoding a major sperm protein containing an ankyrin-repeat domain and an E3-Ubiquitin ligase respectively, and are required for the development of an endosymbiotic-specific structure called the infectosome[49]. The infectosome is a specialized cellular structure formed during AM or the N-fixing root nodule symbiosis, and it is involved in the intracellular colonization of plant cells by symbiotic microbes[49–53].

Based on these observations, it can be hypothesized that transcriptional activation of infectosome-related genes during ErM relies on transcription factors involved in the development of the infectosome during plant endosymbiotic interactions. Indeed, we observed the upregulation of genes encoding putative GRAS and AP2/ERF domains-containing transcription factors in *C. fissa* during ErM that were also transcriptionally responsive in *M. palaecea* during AM*.* For example we detected upregulation of *GRAS17* (Calfis_g10023), and a gene encoding a putative AP2/ERF domain-containing transcription factor (Calfis_g17180, Supplementary Dataset 4). Phylogenetic analysis of the *C. fissa* AP2/ERF encoding gene indicated that it belongs to a clade of duplicated genes in angiosperms that includes transcription factors with established functions during both AM and the root nodule symbiosis, such as *ERN1, ERN2,* and *ERM1* genes (Figure 5C, Supplementary Dataset 4).

The highly specific transcriptional activation pattern of the infectosome-gene module in both ErM and AM reported in this study, including the upstream receptor EPR3*,* the ERN-like transcriptional regulators, and the two minimal infectosome components, LIN and VAPYRIN, suggest that the evolution of plant endosymbiosis, including ErM, required this infectosome-gene module.

The recruitment of infectosome-related genes from AM for the accommodation of alternative plant endosymbioses, however, has been proposed as a key mechanism underlying the diversification of the nodule symbiosis from AM in angiosperms[49,52,54]. Thus, besides the conservation of the minimal infectosome components, LIN and VAPYRIN, endosymbiosis-specific components might have been recruited for the evolution of alternative symbiotic types in land plants. This hypothesis is indeed supported by empirical evidence. For example, the recruitment of RPG proteins to the infectosome, along with VAPYRIN and LIN, facilitates rhizobia entry into roots and has been proposed as a key innovation in the evolution of the nodule symbiosis in legumes[49,55,56]. Additionally, it has been determined that in the context of arbuscule formation, VAPYRIN can specifically interact and mediate the recruitment of the exocyst complex component Exo70I[57]. Interestingly, our analysis identified specific upregulation of a gene encoding the putative exocyst complex component, Exo84, under the ErM-permissive condition in *C. fissa* plants (Supplementary

Dataset 3). We interpret this as an evolutionary novelty, a specific component linked to the intracellular accomodation of the fungal symbiont, that might facilitate the establishment of ErM in host plants.

## Conclusion

In this study, we provide the first genomic resources for leafy liverworts to investigate the evolutionary and molecular processes governing ErM in land plants. By employing genomic and transcriptomic analyses in the resynthesized symbiotic model established here, between *C.fissa* and *H. hepaticicola*, we observed set of genes involved in signaling, and fungal intracellular accommodation during ErM that are also transcriptionally responsive during AM and the root-nodule symbiosis. The discovery that a conserved infectosome gene core module, comprising VAPYRIN and LIN, underlies the intracellular accommodation of completely distinct microbial endophytes in alternative endosymbiotic types may provide valuable insights for guiding efforts to develop synthetic intracellular symbioses in land plants.

## Acknowledgements

We are grateful to Sylvia Perotto (University of Turin) for kindly providing the *Hyaloscypha hepaticicola* (syn. *Rhizoscyphus ericae* UAMH 6735) strain used in this study. We thank the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, https://doi.org/10.15454/1.5572369328961167E12), the NIG supercomputer at ROIS National Institute of Genetics,and the Data Integration and Analysis Facility at the National Institute for Basic Biology for providing computing resources. We acknowledge the TRI-FRAIB imaging facility, member of the national infrastructure France-BioImaging supported by the French National Research Agency (ANR-10-INBS-04). Research performed at LRSV was also supported by the “Laboratoires d” Excellence (LABEX)’ TULIP (ANR-10-LABX-41) and the “École Universitaire de Recherche (EUR)” TULIP-GS (ANR-18-EURE-0019). This work was supported by the project Engineering Nitrogen Symbiosis for Africa (ENSA) funded through a grant to the University of Cambridge by the Bill and Melinda Gates Foundation (OPP1172165) and the UK Foreign, Commonwealth and Development Office as Engineering Nitrogen Symbiosis for Africa (OPP1172165); the Horizon Europe programme (MSCA-PF grant 101105838 ‘SYMBIOLOSS’ to MEB); and the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement no.101001675-ORIGINS) to P.-M.D. MEXT and JSPS Kakenhi JP22128008 and JP23657063 to TN. *S. infuscum* culture to library construction were performed by K. Yamada. Illumina sequencing of *S. infuscum* was performed by Dr. Katushi Yamaguchi and

Dr. Shuji Shigenobu at the Core facility of NIBB through NIBB Collaborative Research Program (12-716) to TN. Celophane was provided by Futamura Chemical, Co., LTD. We thank Dr. Elizabeth Jamet (LRSV, CNRS) for kindly allowing us to use the camera for phenotyping. We are grateful to Dr. M. Shimamura for pointing out the change in taxonomy.

**Data availability**

The *Jungermannia infusca* genome assembly with annotation and raw reads are available on INSDC under the BioProject accession PRJDB19613.

The *Calypogeia fissa* genome assembly and annotation is availablee on ENA-EMBL, with the reference number PRJEB86056. All DEG results, functional annotation, orthogroups and template scripts have been deposited in the repository 10.57745/MD77E8.

An interactive Genome Browser can be found and queried here: https://bbric-pipelines.toulouse.inra.fr/myGenomeBrowser?browse=1&portalname=Calypogeia_fissa&owner=jean.keller@cnrs.fr&key=DKW4edLL.

The raw RNAseq read of *Calypogeia fissa* are available on the SRA database, with the reference number PRJNA1186220.

**Material and methods**

**Preparation of high molecular weight (HMW) DNA for sequencing**

*C. fissa* genomic DNA was isolated from whole plant using QIAGEN Genomic-tips 500/G kit (Cat No./ID: 10262) following the tissue protocol extraction. Briefly, 15g of plant material were frozen and ground in liquid nitrogen with mortar and pestle. After 1h of lysis at 50°C and one centrifugation step, the DNA was immobilized on the column. After several washing steps, DNA is eluted from the column, then desalted and concentrated by alcohol precipitation. The DNA is resuspended in EB buffer. DNA quality and quantity were assessed respectively using the Nanodrop-one spectrophotometer (Thermo Scientific) and the Qbit 3 Fluorometer using the Qbit dsDNA BR assay (Invitrogen). The size of the DNA was assessed using the FemtoPulse system (Agilent, Santa Clara, CA, USA).

**DNA extraction and short reads sequencing**

Genomic DNA from *S. infuscum* was extracted with the CTAB method and purified using a Qiagen genomic Tip 20. A 180-bp and 700-bp insert Illumina paired-end libraries were constructed with a TruSeq PCR-free Sample Prep kit from covaris S2 fragmented and size-fractionated DNA. A 350-bp insert library was constructed with a gel-free protocol. Multiple mate-pair libraries were constructed with a Nextera mate pair library preparation kit. At first,

fragmentation were carried out with 4 µg of DNA and three fractions were recovered after gel electrophoresis. A second mate-pair library set was constructed starting 10 µg of DNA and seven fractions were recovered after CHEF electrophoresis (switch time 1-4 sec, 16 hr, 0.5X TBE, 1% Megabase Agarose). Total RNA was extracted from 6-day old tissue using Qiagen RNA mini kit with RLT buffer. RNA-seq libraries were constructed with a TruSeq Stranded mRNA Prep kit with standard polyA selection or rRNA depletion usingRibo-Zero Magnetic kit (Plant leaf). Sequencing of the libraries were carried out on Illumina sequencers HiSEQ2000/2500 and MiSeq at the National Institute for Basic Biology, Okazaki, Japan.

**Preparation and sequencing of HiFi PacBio library**

A HiFi SMRTbell® library was prepared using the SMRTbell® Template Prep Kit 2.0 from Pacific Biosciences, following the manufacturer's instructions. High-molecular-weight DNA was fragmented to around 20 kb, then treated to remove single-strand overhangs and repair damage. Overhang adapters were attached to both ends of the DNA to form a closed circular DNA molecule, with further clean-up done using the SMRTbell® Enzyme Clean-up Kit 2.0. The library's size and concentration were measured with the FemtoPulse and Qubit systems. Sequencing primer and DNA polymerase were added before loading onto a SMRT cell, and sequencing ran for 30 hours on the Sequel® II system at the University of Lausanne, Switzerland.

**Whole genome assembly**

The *C.fissa* genome was assembled using HiFiasm assembler (v16.1)[58]. Hifiasm is able to produce a primary assembly and alternative assembly (incomplete alternative assembly consisting of haplotigs under heterozygous regions. We then filtered the primary assembly from organelles and low quality contigs. Organelles were identified by mapping contigs on NCBI reference assemblies. The closest references (NC_043787.1 and MF401632.1) were found by MitoHiFi software[59]. We finally removed 1009 contigs from the primary assembly (mitochondrial and chloroplastic contigs identified, high covered (>100X), low covered (<1X) and high %GC (>50%) contigs). All the metrics correspond to the filtered primary assembly used for the next analysis. We obtained 1,396,370 corrected reads (N50 = 17.4kbp; Genome Coverage = 32X) assembled and filtered in 82 contigs (N50 = 15.8Mbp; L50 = 14 contigs; GC content = 42.94%) for a total assembly size of 763Mbp.

To assess the completeness and quality of the genome, we used the Benchmarking Universal Single-Copy Orthologs (BUSCO)[60] tool with the viridiplantae database (n=425). We obtained a 93.7% complete BUSCO score on the assembly.

The *S. infuscum* genome was assembled from Illumina short reads data using the ALLPATHS-LG program R49548 with default parameters.

**Genomes annotation**

The *C. fissa* genome was first softmasked using EarlGrey v4.1.5[61] with default parameters and the NCBI taxid 33090 (Viridiplantae) for the RepeatMasker search. The -d option was set to 'yes' to produce a softmasked version of the genome.

Gene prediction in the *C. fissa* genome was done using BRAKER3.0.7[62] in ETP mode with both RNASeq and protein hints. RNASeq hints were produced from the transcriptomic data generated in this study and described below. Briefly, each sample was first mapped against the *C. fissa* genome using HISAT2 v2.2.1[63] with the following parameters: --sensitive --no-discordant --no-mixed --dta --max-intronlen 5000. Then each mapping file was processed with Samtools v1.19[64] collate, fixmate (with the -m -c and -r options), sort and markdup. Finally, all mapping files were merged into a unique file with the merge and collate options of Samtools prior being processed as described for individual mappings. Both mapping and protein database files were used by BRAKER to annotate the genome using the following tools as contained in the BRAKER 3.0.7 Singularity container (available at: https://hub.docker.com/r/teambraker/braker3): miniprot[65], GeneMark-ETP[66], diamond[58,], Spaln2[59], StringTie2[69], GFF utilities[70], AUGUSTUS[71,72] and TSEBRA[73]. To maximize the completion of the prediction, BRAKER was run with compleasm[74] option set to "Viridiplantae".

BRAKER prediction was then processed to remove the alternative transcript using agat_sp_keep_longest_isoform.pl from the AGAT v1.4.0 toolkit[75] and coding-sequence as well proteins extracted from the genome using gffread v0.12.8 from GFF utilities package[70] with the -C and -V options enabled. Proteins were functionally annotated uysing InterproScan5 v64.96.0[76] with the options '-iprlookup' and 'goterms' enabled.

The *S. infuscum* genome has been annotated using RepeatModeler Version open-1.0.7, RepeatMasker version open-4.0.5 (NCBI/RMBLAST 2.2.27+), exonerate (2.2.0), bowtie2 (perhaps 2.1.0), tophat2 (2.0.10), bowtie (1.0.0), and AUGUSTUS v. 3.0.2 using scripts available at https://github.com/tomoakin/Jin-annot. Arabidopsis (TAIR10) and *Physcomitrium patens* (*P.patens*.V6_filtered_cosmoss_proteins.fas) protein datasets were used for exonerate mapping. 224 loci were manually checked on Web Apollo[77] and converted to genbank format for training Augustus, *Marchantia polymorpha* proteins collected from GenBank as of 1 Feb, 2017 were additionally mapped to the reference with exonerate. RNA-seq data were remapped with hisat2, and the final run using with UTR=on (version 3.2.3).

**Whole genome duplication analysis in *C. fissa***

Paralogous Ks was calculated using the wgd package v1.1.2[78]. Specifically, genes were clustered with the mcl command using an inflation index of 2.0 prior to estimating paralogous

Ks with the ksd command using the --pairwise flag. Orthologous Ks was estimated between *C. fissa* and *J. infusca* in a similar manner using the mcl command using the --one_v_one and with the ksd flag under default settings.

Syntenic gene pairs were identified and plotted using the python version of MCSCAN as implemented in the jcvi comparative genomics toolkit v0.9.2[79]. First, a file containing CDS sequences was generated with gffread[70] using the -x flag and the annotation file was converted from gff3 to bed format using the 'jcvi.formats.gff bed' function for *C. fissa* and *M. polymorpha*. The resulting CDS and bed files were used as input 'jcvi.compara.catalog ortholog' function to identify and visualize synteny within and between species. A default C-score of 0.7 was used to filter low quality hits.

**Plant and fungal material, growth conditions, inoculation assays, and nutrient supply experiment**

*Calypogeia fissa* (L.) Raddi ssp. *fissa* was sourced from the Edimburgh botanical garden and maintained in polypropylene microboxes (OV80 + OVD80L#10G, Sac O2, Dutscher no. 017055, Issy-les-Moulineaux, FR) with 150 g of silica sand (Sibelco, Puel, no. EMB33284, Toulouse, FR). Microboxes were supplemented with 40 mL Long-Ashton solution (standard P, high N; see Supplementary Table 1) and incubated at 23 °C, 16-hour day/ 8-night at 20 °C cycle. For the time-series experiments, plants were watered at weeks 4 and 7 post-inoculation. Permissive and non-permissive nutrient solutions for ericoid mycorrhizae in *C. fissa* were selected by screening the effects of seven different recipes with varying phosphate and nitrogen concentrations, based on the Long Ashton nutrient solution[80]. The selection was made by assessing plant and fungal growth, as well as colonization levels of *C. fissa* rhizoid tips by *H. hepaticicola*. Recipes for the nutrient solutions used in this study can be found in Supplementary Table 1.

*Hyaloscypha hepaticicola* (syn. *Rhizoscyphus ericae* UMAH 7357) (D.J. Read) Zhuang & Korf was maintained via mycelial agar plug transfers on Czapek-Dox (C-Dox) agar plates prepared per Atlas[81]. Before *C. fissa* cultivation, three agar plugs were transferred to microboxes containing silica sand and nutrient solution and incubated in dark conditions at 17°C for two weeks. Subsequently, ten *C. fissa* stems were introduced into each microbox, which further developed into plant colonies. For nutrient supply studies on plates, *R. ericae* agar plugs were added to plates with a nutrient medium consisting of working solutions from Supplementary Table 1, supplemented with 3.5% sucrose and 1.5% BD Bacto™ Dehydrated Agar (Fisher Scientific, no. 214010). Plates were incubated as described above.

*Solenostoma infuscum* (Mitt.) Hentschel (*syn J. infusca*) has been maintained as axenic culture at Kumamoto University and provided by Prof. Hiroyoshi Takano. The plants were

grown on KNOP agar plates covered with a sheet of cellophane (provided by Futamura Chemical, Co., LTD.) after cutting with a Polytron similar to *Physcomitrium patens* culture but shorter time of cutting than protonema culture.

**Phenotyping**

Imaging of microboxes and fungal plates was conducted using a commercially available Canon EOS 4000D digital single-lens reflex camera (18.0 megapixels). Weekly images of microboxes containing three *C. fissa* colonies were captured to monitor growth. Colony area was estimated using ImageJ software[82]. The scale was set using the known width of the microbox (12.5 cm), and color thresholds were adjusted to optimize hue (35–30 upper, 255 lower), saturation (51 upper, 255 lower), and brightness (0 upper, 255 lower) for each image. Particle analysis was then performed with parameters set to a size range of 10,000–Infinity $cm^2$, circularity of 0.00–1.00, and results displayed as outlines. Outputs, including the processed image, analyzed data, and measurement summary, were saved for further analysis. The colony area ($cm^2$) measured weekly is reported as a proxy for biomass change in this study.

**Microscopy and quantification of colonization events**

Ten independent stems from three independent microbox biological replicates from inoculated and non-inoculated *C. fissa* samples were prepared for microscopy using Wheat Germ Agglutinin (WGA-cf488a) and Congo Red to visualize chitinous structures and cell walls. Samples were fixed in 10% (v/v) at room temperature overnight, washed three times in water, and resuspended in 1X (v/v) phosphate-buffered saline (PBS), containing 1 µg/mL WGA (Dominique Dutscher, no. 462396, Issy-les-Moulineaux, FR), and incubated overnight at room temperature in the dark. For Congo Red staining, samples were incubated in 0.5% congo red aqueus solution (Rouge Congo pur RAL Diagnostics, LABELIANS, no. 3634100005, Nemours, FR) for 10 minutes just before confocal acquisitions. After staining, samples were rinsed twice in water to remove excess dye and mounted on microscope slides in water to be observed.

Microscopy was performed using both epifluorescence and confocal microscopy. Epifluorescence imaging was conducted using a Zeiss Axiozoom V16 equipped with objective 0.5x, GFP like filter set (Zeiss 38 He filter set, ex:470/40, em: 525/50) and RFP filter set (Zeiss 20 filter set, ex:546/12, em: 575-640) for WGA and Congo Red visualization, respectively. We also used Nikon Eclipse Ti wide field microscope with Nikon GFP filter set (ex: 472/30, em: 520/35). Confocal microscopy was carried out using a Leica SP8 CSU and objective 25x/0.95 water immersion. WGA was excited at 488nm and emission set between

500 and 545 nm. Congo Red was excited at 552nm and fluorescence was collected between 590 and 640 nm. Z-stack was performed with step size range between 0.57µm to 1µm.

Image analysis and processing were performed using ImageJ and LAS X Life Science Microscope Software, with identical settings applied across all samples to ensure consistency.

Manual counting of *C. fissa* swollen rhizoid tips was performed using a four-digit manual cell counter (4.3 × 3.5 mm with reset knob, Avantor VWR no. MOIN1494MW, Rosny-sous-Bois, FR) under a Zeiss Axio Zoom.V16 microscope. Results obtained from the Axio Zoom.V16 were independently validated using a Nikon Eclipse Ti microscope.

**Harvesting of plants and RNA extraction**

Whole *C. fissa* colonies were harvested during the morning on week 8 post-inoculation. Plants were pooled from three replicate microboxes (three independent colonies with hundreds of plant individuals per colony) per experiment, frozen in liquid nitrogen and stored in 50-mL screw-cap polypropylene tubes at −80°C until further processing. Frozen plant tissues were homogenized in 50-ml polypropylene tubes (Avantor Performance Materials GmbH, Griesheim, Germany, Catalog no. 734-0453) pre-chilled in liquid N2, each containing a commercial glass marble, using a vortex (Vortex-Genie 2, Scientific Industries, Inc., Bohemia, New York, USA) at maximal speed for 10 s, with three to four cycles of grinding and chilling in liquid N2. RNA was isolated from aliquots of 300 mg of frozen homogenized tissue using the Direct-zol RNA Microprep Kit following supplier's instructions (Zymo Research Europe GmbH, Catalog no. R2062, Freiburg, DE).

**Estimation of differentially expressed genes**

RNASeq raw reads from all described conditions were mapped against the genome of *C. fissa* and counted using the Nextflow v23.10.0[83] pipeline NF-CORE/RNASeq v3.14[84], the options star_salmon to align and quantify reads as well as '-nextseq 30 -length 50' as extra parameters of TrimGalore v0.6.7 (10.5281/zenodo.5127899) to remove reads with quality lower than 30 or a length lower than 50bp. Ribosomal RNA were also removed through the option "-remove_ribo" using SortMeRNA v4.3.4[85]. The pipeline was run under the GenoToul configuration (https://github.com/nf-core/configs/blob/master/docs/genotoul.md) and used the following softwares and languages: bedtools v2.30.0, R v4.0.3, v4.1.1 and v4.2.1[86], DESEQ2 v1.28.0[87]; dupradar v1.28.0[88], fastqc v 0.12.1 (bioinformatics.babraham.ac.uk/projects/fastqc/), fq v0.9.1 (https://github.com/stjude-rust-labs/fq), gffread v0.12.1[70], perl v5.26.2 (dev.perl.org/perl5/news/), python v3.9.5 (python.org/downloads/release/python-395/), rsem v1.3.1[89], STAR v2.7.10a[90], picard v3.0.0 (broadinstitute.github.io/picard), qualimap v2.3[91], rseqc v5.02[92], salmon v1.10.1[93],

summarizedExperiment v1.24.0[94], samtools v1.16.1[64], stringtie v2.2.1[69], tximeta v1.12.0[95], UCSC v377 and v445 (https://github.com/ucscGenomeBrowser/kent).

For each condition and time point, differentially expressed genes (DEGs) were estimated independently with edgeR v4.2.1[96] in R v4.4.0 using the quasi-likelihood F-tests method and a FDR threshold of 0.05.

Enriched functions were determined using the R package clusterProfiler v4.12.3[97], the DEGs with a logFC threshold of |1| as foreground and genes subjected DEG analysis (*i.e.,* genes that were retained after removing the low counts) as background.

**Orthogroups reconstruction**

Orthogroups at the scale of Bryophytes were reconstructed using OrthoFinder v2.5.5[98] covering sequenced Bryophyte lineages. Prior to orthogroup reconstruction, genome annotation of each species was processed as described for *C. fissa*: retention of longest isoforms with AGAT and then extraction of sequences with gffread and the "-C -V" options. OrthoFinder was then run with the "diamond_ultra_sens" and the msa options[99]. For all downstream orthogroup-based analyses presented in this work, the phylogenetic hierarchical orthogroups containing all species (N0) has been used.

**Cross-referencing of transcriptomic data and orthogroups**

Orthogroups were then filtered to retain the ones containing only up-regulated genes of *C. fissa* in the ErM-permissive condition, no differential expressed genes in the restrictive condition and at least one gene up-regulated in *M. paleacea* at 3, 5 or 8 week post-inoculation.

**Ortholog inference of symbiosis-related genes**

We reconstructed the evolution of 16 gene families containing genes with a known role in symbiosis. We used BLAST v2.15.0[100] to search for putative homologs in the genomes of *C. fissa* and another 25 species, including 15 bryophytes and 10 vascular plants, with the protein sequences of *Marchantia paleacea* and *Medicago truncatula* as queries (*e-value*= $10^{-9}$). Sequences were aligned using MAFFT v7.505[101] with the automatic mode (*--auto*, *--maxiterate* 20), and alignments were subsequently trimmed using trimAl v1.4.1[102] to remove columns with more than 70% of gaps (*-gt* 0.3). An approximately maximum likelihood tree was then estimated using FastTree v2.1.11[103] with the gamma model (*-gamma*). Trees were inspected to identify the group of orthologous sequences containing the query, which was extracted, realigned using MUSCLE v5.1.0[104] and trimmed with trimAl to remove columns with more than 90% of gaps (*-gt* 0.1). A maximum-likelihood tree was then estimated with IQ-TREE v2.2.2.6[105] using the standard model selection (*-m TEST*), with branch support

assessed using the SH-like approximate likelihood test and ultrafast bootstrap, both with 1000 replicates.

**Selection tests on arbuscular mycorrhizal symbiosis-related genes**

We used the framework of branch models[106,107] to test for shifts in selection along the evolution of Jungermanniales on three protein-coding genes associated with arbuscular mycorrhizal symbiosis (RAD1, STR and STR2). Protein alignments of bryophyte species were retrieved from the curated sets of orthologs of symbiosis-related genes (see above) and used to align the corresponding nucleotide sequences as codons using PAL2NAL v14[108]. Alignments were trimmed using trimAl to remove all positions containing gaps (*-gt* = 1), and phylogenetic trees were estimated using IQ-TREE with the standard model selection. The resulting codon alignment and gene tree were used for the optimization of branch models using the program codeml as part of PAML v4.10.7[109]. Here we compared the null model, which assumes a single ratio of nonsynonymous to synonymous substitutions rates (dN/dS = ω) for all branches in the tree, to an alternative model in which we hypothesized a different ω for the Jungermanniales branches (the *Calypogeia* + *Jungermannia* group, including its internal branch) relative to the other branches in the tree. A likelihood ratio test (LRT) was used to assess the performance of the alternative model compared to the null at a significance level of 5%.

**Estimation of divergence times**

Divergence times among bryophyte lineages were estimated using a set of nuclear orthologous genes extracted from the orthogroup data set (see above). Because no single-copy orthogroups were detected, we selected those with duplicated genes in only one species (total = 346 orthogroups), in which case only the longest of the duplicates was retained. Orthogroups containing putative pseudogenes or non-homologous sequences were then identified and removed from the data set. For that, nucleotide sequences were codon-aligned as described above, and phylogenetic trees were inferred using IQ-TREE with the standard model selection. We then used TreeShrink v1.3.9[110] to identify orthogroup trees containing outlier long branches with a false positive rate of 5% ($q$ = 0.05), and removed these orthogroups from the data set. Finally, we randomly selected 50 orthogroup alignments, which were concatenated and trimmed using trimAl to remove all positions containing gaps (*-gt* = 1; final alignment length = 60,594 bp). A time-calibrated phylogenetic tree was estimated using BEAST v2.7.4[111]. We used the GTR+G substitution model with four gamma categories, with the Yule model as speciation prior and an optimised relaxed clock[112]. We adopted here a secondary calibration strategy using divergence time estimates from published data[23], which represents, to our knowledge, the most comprehensive

assessment to date of the fossil record of land plants for node calibration of molecular dating. We fixed the age of the root at 464.5 Ma, and the age of crown mosses at 392 Ma, both ages corresponding to the respective mean values from Harris et al., and that were used here as the mean of a normal distribution with standard deviation of 0.001. To incorporate the current understanding of the phylogeny of bryophytes and reduce tree search space, we fixed the position of *Takakia lepidozioides* as sister to the other mosses[113] and the monophyly of the three main bryophyte lineages[23]. Two MCMC chains were run in parallel for 400,000,000 generations each, sampling every 10,000 generations. We monitored the runs using Tracer v1.7.1[114] to check for convergence and effective sample sizes > 200 for all parameters. We discarded trees sampled before the convergence of the runs (burn-in = 10%) and combined all the remaining trees into a maximum clade credibility tree, with the median ages summarized on the nodes.

### Nicotiana benthamiana transactivation assays

The Golden Gate modular cloning system was used to prepare the plasmids as previously published[12,115,116]. Levels 0, 1, and 2 constructs used in these assays are listed in Supplementary Dataset 5 and held for distribution in the ENSA project core collection (https://www.ensa.ac.uk/). *CsCYCLOPSa*, *CsCYCLOPSb*, *NmCYCLOPS* and *MtCYCLOPS* sequences were domesticated, synthesized, and cloned into pMS (GeneArt, Thermo Fisher Scientific, Waltham). *MtCCaMK-Kin* and *CYC-RE$_{pro}$* sequences were taken from published data[35].

Constructs were introduced into *Agrobacterium tumefaciens* GV3101 and co-infiltrated with P19 into *N. benthamiana* leaves as described elsewhere[117]. For co-infiltration, equal volumes of each *A. tumefaciens* different cultures were adjusted to an $OD_{600}$ of 0.5 and mixed prior to infiltration. Each combination of constructs were infiltrated into three different leaves of two different plants. 48 hours *post* inoculation, leaf discs were harvested and GUS stained as previously described[35] with the following modifications: 0.1% triton was added to the GUS solution and leaf discs were first placed under vacuum for 20 min in the GUS solution before incubation at 37°C for 20 hours. Discs were rinsed with water and treated with ethanol 95% to remove the chlorophyll.

## Figure legends

**Figure 1. A recent whole-genome duplication in the leafy liverwort *Calypogeia fissa*.**

**(A)** Distribution of synonymous substitutions per site (Ks) between paralog pairs and syntelogs in the *C. fissa* genome. **(B)** Intragenomic synteny analysis identifies conserved collinear gene arrangements

within the *C. fissa* genome, revealing 137 collinear blocks comprising 7,087 syntelogs and covering 37.2% of the *C. fissa* genome. **(C)** Synteny between *C. fissa* and *Marchantia polymorpha*. **(D)** Time-calibrated phylogeny of bryophyte species. Node bars represent the 95% highest posterior density (HPD) intervals. The root of the tree and the crown node of mosses were fixed to the mean ages (see Methods). Circles on the right are proportional to genome sizes and the number of coding sequences (CDS) in the genomes of the corresponding species.

**Figure 2. The establishment of ericoid mycorrhizae in *Calypogeia fissa* is maximized under conditions of both phosphate and nitrogen deficiency. (A-F)** Representative images of *C. fissa* plants growing in mock-inoculated (non-inoc) and inoculated (inoc with *Hyaloscypha hepaticicola*) treatments, as indicated, under phosphate (P) and nitrogen (N) deficiency conditions (A, B), or water (D, E) (see Suppl. Table 1). Scale bars: 2 cm. Estimated growth of *C. fissa* in non-inoc and inoc treatments supplied with a low P/N solution (E) or water only (F). Data points represent the mean ± SEM ($n$ = 27) of technical replicates per time-point from one representative experiment of a total of three independent experiments. **(G-M)** Representative fluorescence microscopy photographs of *C. fissa* stems harvested from non-inco and inoc treatments and stained with Wheat Germ Agglutinin (WGA). Plants were photograph and harvested after 9 weeks of growth. Scale bars: 200 µm. **(N-O)** Confocal microscopy images of samples shown in (M), and further stained with Congo red (blue-fluorescent signal), showing the development of swollen-tip rhizoids where *H. hepaticicola* is intracellularly accommodated (green-fluorescent signal). Scale bars: 200 µm (N), 75 µm (O), and 50 µm (I). **(P)** Total number of swollen and colonized rhizoids per one stem in non-inoc and inoc treatments as indicated in (see also E, F). Data is based on individual quantification of intracellular colonized rhizoids stained with WGA under epifluorescence and confocal microscopes. Data points represent the mean ± SD ($n$ = 4 to 11) of individual technical stem replicates from one representative experiment. In (E, F), different characters indicate statistically significant differences based on a Welch's (double-tailed) $t$-test ($P < 0.05$). See also Supplementary Tables 3.

**Figure 3. The establishment of ericoid mycorrhizae in *Calypogeia fissa* is promoted under phosphate limiting conditions and is restricted by high phosphate and moderate N supply.**
**(A-F)** Representative images of *C. fissa* plants growing in mock-inoculated (non-inoc) and inoculated (inoc *with* Hyaloscypha hepaticicola) treatments, as indicated, under conditions of high phosphate (P) and moderate nitrogen (N) supply (A, B), or low P and moderate N supply (D, E) (see Supplementary Table 1). Scale bars: 2 cm. Estimated growth of *C. fissa* in non-inoculated and inoculated treatments under the indicated nutrient conditions (C, F). Data points represent the mean ± SEM ($n$ = 27) of technical replicates per time-point from one representative experiment of a total of three independent experiments. **(G-L)** Representative fluorescence microscopy photographs of *C. fissa* stems harvested from mock-inoculated and inoculated treatments and stained with Wheat Germ Agglutinin (WGA, see Methods). Plants were photograph and harvested after 8 weeks of growth (Methods). Scale bars: 200 µm (G, J). Confocal microscopy images of samples shown in (K, L), were further stained with Congo red (blue-fluorescent signal), and show the development of swollen-tip rhizoids where H. hepaticicola

is intracellularly accommodated (green-fluorescent signal). Scale bar in (L): 25 µm. **(M)** Total number of swollen and colonized rhizoid tips per one stem in non-inoc and inoc treatments as indicated in (C, F). Data is based on individual quantification of intracellular colonized rhizoids stained with WGA under epifluorescence and confocal microscopes. Data points represent the mean ± SD ($n$ = 4 to 13) of technical stem replicates from one representative experiment. In (C, F), different characters indicate statistically significant differences between inoculated and non-inoculated treatments, based on a Welch's (double-tailed) $t$-test ($P < 0.05$) (C) or based on a Kruskal-Wallis test with post-hoc Dunn's test. See also Supplementary Tables 3.

**Figure 4. Conservation and functional validation of ortholog endosymbiotic-related genes in ericoid mycorrhizae-forming leafy liverworts. (A)** Conservation of key genes with known roles during mycorrhizae in the genomes of bryophyte species. **(B)** Conserved transactivation of the *CYC-REpro:GUS* reporter construct by the transcription factor CYCLOPS in Jungermanniales. *Nicotiana benthamiana* leaves were infiltrated with the *CYC-REpro:GUS* construct alone, or with an overexpressing *MtCCaMK* construct in combination with overexpressor constructs of different *CYCLOPS* orthologs from *Medicago truncatula* (*Mt*), *Calypogeia fissa* (*Cf*), or *Scapania nemorosa* (*Nm*). Leaves were incubated for 48h with the GUS-staining solution and cleared with ethanol. One representative leaf disc out of six discs is shown for each combination.

**Figure 5. Transcriptional reprogramming in *Calypogeia fissa* during ericoid mycorrhizae with *Hyaloscypha hepaticicola*. (A)** Differential gene expression analysis comparing mock-treated and *H. hepaticicola*-inoculated *C. fissa* plants under ericoid mycorrhiza-permissive conditions. A total of 787 genes were upregulated, and 743 genes were downregulated in inoculated plants relative to mock-treated controls. **(B)** Differential gene expression analysis under ericoid mycorrhiza-permissive restrictive conditions. In this scenario, 566 genes were upregulated, and 27 genes were downregulated in inoculated plants compared to mock-treated controls. **(C)** Comparative transcriptomic analysis of ericoid mycorrhiza-permissive and restrictive conditions alongside arbuscular mycorrhiza-permissive conditions. The orange-overlapping area in the Venn diagram depicts the conserved expressed gene cluster that was found to be upregulated during both ericoid and arbuscular mycorrhizae. **(D)** Expression bar plots displaying four well-known genes with key functions in endosymbiotic interactions, all of which were found to be upregulated in both ericoid and mycorrhizal symbiosis under permissive conditions. Cross-comparison analysis were done as described in the Methods section.

## Tables

**Table 1. Genome assembly statistics for the leafy liverworts *Solenostoma infuscum* and *Calypogeia fissa*.**

## Supplementary information

**Supplementary Figure 1. Orthologous Ks distribution in Calypogeia fissa and Solenostoma infuscum provides evidence for a recent Whole Genome Duplication (WGD) in C. fissa.**

The Ks (synonymous substitution rate) distribution of orthologous genes between C. fissa and *S. infuscum* was assessed to investigate the timing of whole genome duplication (WGD) events. The Ks plot shows a prominent peak around Ks = 0.35, indicating a recent WGD event in *C. fissa*. This peak is consistent with a scenario in which WGD occurred after the divergence between C. fissa *and S. infuscum*. The distribution suggests that the recent WGD in *C. fissa* predates the split between these two species.

**Supplementary Figure 2. Intergenomic synteny comparison between Calypogeia fissa and Marchantia polymorpha reveals limited collinearity.** An intergenomic comparison of synteny between *C. fissa* and the distantly related species *M. polymorpha* was conducted to assess the conservation of gene order. The analysis identified relatively few collinear gene pairs, with approximately 10% of genes in each species showing synteny. This limited collinearity is possibly due to the deep phylogenetic divergence between the two species.

**Supplementary Figure 3. Growth of *Hyaloscypha hepaticicola* on agar plates supplemented with variable phosphate and nitrogen nutrient regimes. (A-G)** Representative images of *H. hepaticicola* grown under the indicated nutrient conditions (A-F) used for physiological characterization of *C. fissa* (see Supplementary Table 1) or a basal medium consisting of water, 1.5% (w/v) sucrose, and 0.8% (w/v) Bacto agar (see Methods for details). Plates were incubated at 17°C in darkness and photographed at week 23.

**Supplementary Figure 4. Establishment of ericoid mycorrhizae in *Calypogeia fissa* under phosphate limiting conditions. (A-B)** Representative images of *C. fissa* growing in mock-inoculated (non-inoc) and inoculated as indicated, under conditions of low phosphate (P) and high Nitrogen (N) supply (see Figs. 2 and 3). Scale bars: 2 cm. **(C)** Estimated growth of *C. fissa* growing in non-inoc and inoc treatments, in low P conditions as indicated. Data points represent the mean ± SEM ($n$ = 27) of technical replicates from one representative experiment of a total of three independent experiments. **(D-E)** Representative fluorescence microscopy photographs of *C. fissa* stems harvested from non-inoc and inoc treatments and stained with Wheat Germ Agglutinin (WGA, see Methods). Plants were photograph and harvested after 9 weeks of growth. Scale bars: 200 µm. **(F)** Total number of swollen and colonized rhizoid tips per one stem in non-inoc and inoc treatments as indicated in (C). Data points represent the mean ± SD ($n$ = 7 to 13) of technical stem replicates from one representative experiment. In (C, F), different characters indicate statistically significant differences based on a Welch's (double-tailed) $t$-test ($P < 0.05$). $P$ values in (C) are based on a Kruskal-Wallis test with a post-hoc Dunn's test. **(G, J)** Confocal microscopy images stained with WGA and Congo red (blue-fluorescent signal), showing intracellular colonization by *H. hepaticicola* (green-fluorescent signal) and the development of swollen rhizoid tips in *C. fissa* stems. Scale bar in (G): 200 µm, (H-J): 25 µm.

**Supplementary Figure 5. The establishment of ericoid mycorrhizae in *Calypogeia fissa* is restricted under conditions of high phosphate and limited N supply. (A-F)** Representative images of *C. fissa* growing in mock-inoculated (non-inoc) and inoculated (inoc) treatments, as indicated, under conditions of high phosphate (P) and low N supply (D, E) (see also Figs. 2 and 3). Scale bars: 2 cm. Estimated growth of *C. fissa* in non-inoc and inoc treatments under the indicated nutrient conditions (C-F). Data points represent the mean ± SEM ($n$ = 12) of technical replicates from one representative experiment. **(G-L)** Representative fluorescence microscopy photographs of *C. fissa* stems harvested from non-inoculated and inoculated treatments and stained with WGA. Plants were photographed and harvested after 8 weeks of growth (Methods). Scale bars: 200 µm (G-J). **(M)** Total number of swollen and colonized rhizoid tips per one stem in non-inoc and inoc treatments as indicated in (C, F). Data is based on individual quantification of intracellular colonized rhizoids stained with WGA. Data points represent the mean ± SD ($n$ = 4 to 8) of technical stem replicates from one representative experiment. In (C, F), different characters indicate statistically significant differences between inoculated and non-inoculated treatments, based on a Welch's (double-tailed) $t$-test ($P$ < 0.05).

**Supplementary Figure 6. Maximum-likelihood phylogenetic trees of selected endosymbiosis-related genes based on protein sequences.** Coloured labels indicate high confidence orthologs (blue) and related homologs (yellow) of *Calypogeia fissa,* and the known orthologs of *Marchantia paleacea* and/or *Medicago truncatula* used as query in the analyses (green). Closed circles on branches indicate SH-like support values > 0.8. Highly duplicated taxa were pruned out from the tree to facilitate visualization, and are indicated as triangles at the tip of branches. Scale bars are in expected substitutions per site. Species abbreviations: Ambtri = *Amborella trichopoda* (angiosperms), Adicap = *Adiantum capillus-veneris* (monilophytes), AntagrOXF = *Anthoceros agrestis* cv. OXF (hornworts), Antang = *Anthoceros angustus* (hornworts), Antpun = *Anthoceros punctatus* (hornworts), Aratha = *Arabidopsis thaliana* (angiosperms), Azofil = *Azolla filiculoides* (monilophytes), Calfis = *Calypogeia fissa* (liverworts), Cerpur = *Ceratodon purpureus* R40 (mosses), Cycpan = *Cycas panzhihuaensis* (gymnosperms), Entsed = *Entodon seductrix* (mosses), Fonant = *Fontinalis antipyretica* (mosses), Hypcur = *Hypnum curvifolium* (mosses), Isosin = *Isoetes sinensis* (lycopods), Juninf = *Jungermannia infusca* (liverworts), Luncru64630 = *Lunularia cruciata* (liverworts), Marpal = *Marchantia paleacea* (liverworts), Marpolmon = *Marchantia polymorpha* ssp. *montivagans* SA2 (liverworts), Marpolpol = *Marchantia polymorpha* ssp. *polymorpha* BR5 (liverworts), Marpolrud = *Marchantia polymorpha* ssp. *ruderalis* TAK1 (liverworts), Medtru = *Medicago truncatula* (angiosperms), Orysat = *Oryza sativa* (angiosperms), Phypat = *Physcomitrium patens* (mosses), Psemen = *Pseudotsuga menziesii* (gymnosperms), Ricnat = *Ricciocarpos natans* (liverworts), Selmoe = *Selaginella moellendorffii* (lycopods), Sphfal = *Sphagnum fallax* (mosses), Taklep = *Takakia lepidozioides* (mosses).

**Supplementary Table 1. Chemical composition of macronutrient solutions used in this study**

**Supplementary Table 2. Relative growth response in inoculated and non-inoculated *Calypogeia fissa* plants under varying nutrient conditions across time**

**Supplementary Table 3. Mean number of swollen and colonized rhizoid tips per stem of *Calypogeia fissa* plants with statistical comparisons across nutrient conditions**

**Supplementary Table 4. Branch model tests on arbuscular mycorrhizal symbiosis-related genes**

**Supplementary Dataset 1. Differentially expressed genes in inoculated vs non-inoculated treatments under ericoid mycorrhiza-permissive conditions.**

**Supplementary Dataset 2. Differentially expressed genes in inoculated vs non-inoculated treatments under ericoid mycorrhiza-restrictive conditions**

**Supplementary Dataset 3. Differentially expressed genes in ericoid mycorrhiza-permissive and restrictive conditions**

**Supplementary Dataset 4. Comparative transcriptomic analysis of ericoid and arbuscular mycorrhizae**

## References


1. Taylor, T. N. & Osborn, J. M. The importance of fungi in shaping the paleoecosystem. *Review of Palaeobotany and Palynology* **90,** 249–262 (1996).
2. Brundrett, M. C. Coevolution of roots and mycorrhizas of land plants. *New Phytologist* **154,** 275–304 (2002).
3. Brundrett, M. C. & Tedersoo, L. Evolutionary history of mycorrhizal symbioses and global host plant diversity. *New Phytologist* **220,** 1108–1115 (2018).
4. Cullings, K. W. Single phylogenetic origin of ericoid mycorrhizae within the Ericaceae. *Can. J. Bot.* **74,** 1896–1909 (1996).
5. Schwery, O., Onstein, R. E., Bouchenak-Khelladi, Y., Xing, Y., Carter, R. J. & Linder, H. P. As old as the mountains: the radiations of the Ericaceae. *New Phytologist* **207,** 355–367 (2015).
6. Ward, E. B., Duguid, M. C., Kuebbing, S. E., Lendemer, J. C. & Bradford, M. A. The functional role of ericoid mycorrhizal plants and fungi on carbon and nitrogen dynamics in forests. *New Phytologist* **235,** 1701–1718 (2022).

7. Redecker, D., Kodner, R. & Graham, L. E. Glomalean Fungi from the Ordovician. *Science* **289,** 1920–21 (2000).

8. Krings, M., Taylor, T. N., Hass, H., Kerp, H., Dotzler, N. & Hermsen, E. J. Fungal endophytes in a 400-million-yr-old land plant: infection pathways, spatial distribution, and host responses. *New Phytologist* **174,** 648–657 (2007).

9. Wellman, C. H. The nature and evolutionary relationships of the earliest land plants. *New Phytologist* **202,** 1–3 (2014).

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

11. Radhakrishnan, G. V., Keller, J., Rich, M. K., Vernié, T., Mbadinga Mbadinga, D. L., Vigneron, N., Cottret, L., Clemente, H. S., Libourel, C., Cheema, J., Linde, A.-M., Eklund, D. M., Cheng, S., Wong, G. K. S., Lagercrantz, U., Li, F.-W., Oldroyd, G. E. D. & Delaux, P.-M. An ancestral signalling pathway is conserved in intracellular symbioses-forming plant lineages. *Nat. Plants* **6,** 280–289 (2020).

12. Rich, M. K., Vigneron, N., Libourel, C., Keller, J., Xue, L., Hajheidari, M., Radhakrishnan, G. V., Le Ru, A., Diop, S. I., Potente, G., Conti, E., Duijsings, D., Batut, A., Le Faouder, P., Kodama, K., Kyozuka, J., Sallet, E., Bécard, G., Rodriguez-Franco, M., Ott, T., Bertrand-Michel, J., Oldroyd, G. E. D., Szövényi, P., Bucher, M. & Delaux, P.-M. Lipid exchanges drove the evolution of mutualism during plant terrestrialization. *Science* **372,** 864–868 (2021).

13. Kowal, J., Pressel, S., Duckett, J. G., Bidartondo, M. I. & Field, K. J. From rhizoids to roots? Experimental evidence of mutualism between liverworts and ascomycete fungi. *Annals of Botany* **121,** 221–227 (2018).

14. Pressel, S., Bidartondo, M. I., Field, K. J. & Duckett, J. G. Advances in understanding of mycorrhizal-like associations in bryophytes. *BDE* **43,** (2021).

15. 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 Phytologist* **209,** 1734–1746 (2016).

16. Bowman, J. L., Kohchi, T., Yamato, K. T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., Berger, F., Adam, C., Aki, S. S., Althoff, F., Araki, T., Arteaga-Vazquez, M. A., Balasubrmanian, S., Barry, K., Bauer, D., Boehm, C. R., Briginshaw, L., Caballero-Perez, J., Catarino, B., Chen, F., Chiyoda, S., Chovatia, M., Davies, K. M., Delmans, M., Demura, T., Dierschke, T., Dolan, L., Dorantes-Acosta, A. E., Eklund, D. M., Florent, S. N., Flores-Sandoval, E., Fujiyama, A., Fukuzawa, H., Galik, B., Grimanelli, D., Grimwood, J., Grossniklaus, U., Hamada, T., Haseloff, J., Hetherington, A. J., Higo, A., Hirakawa, Y., Hundley, H. N., Ikeda, Y., Inoue, K., Inoue, S., Ishida, S., Jia, Q., Kakita, M., Kanazawa, T., Kawai, Y., Kawashima, T., Kennedy, M., Kinose, K., Kinoshita, T., Kohara, Y., Koide, E., Komatsu, K., Kopischke, S., Kubo, M., Kyozuka, J., Lagercrantz, U., Lin, S.-S., Lindquist, E., Lipzen, A. M., Lu, C.-W., De Luna, E., Martienssen, R. A., Minamino, N., Mizutani, M.,

Mizutani, M., Mochizuki, N., Monte, I., Mosher, R., Nagasaki, H., Nakagami, H., Naramoto, S., Nishitani, K., Ohtani, M., Okamoto, T., Okumura, M., Phillips, J., Pollak, B., Reinders, A., Rövekamp, M., Sano, R., Sawa, S., Schmid, M. W., Shirakawa, M., Solano, R., Spunde, A., Suetsugu, N., Sugano, S., Sugiyama, A., Sun, R., Suzuki, Y., Takenaka, M., Takezawa, D., Tomogane, H., Tsuzuki, M., Ueda, T., Umeda, M., Ward, J. M., Watanabe, Y., Yazaki, K., Yokoyama, R., Yoshitake, Y., Yotsui, I., Zachgo, S. & Schmutz, J. Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome. *Cell* **171,** 287-304.e15 (2017).

17. Linde, A.-M., Sawangproh, W., Cronberg, N., Szövényi, P. & Lagercrantz, U. Evolutionary History of the Marchantia polymorpha Complex. *Front. Plant Sci.* **11,** 829 (2020).

18. Beaulieu, C., Libourel, C., Mbadinga Zamar, D. L., El Mahboubi, K., Hoey, D. J., Greiff, G. R. L., Keller, J., Girou, C., San Clemente, H., Diop, I., Amblard, E., Castel, B., Théron, A., Cauet, S., Rodde, N., Zachgo, S., Halpape, W., Meierhenrich, A., Laker, B., Bräutigam, A., The SLCU Outreach Consortium, Hoey, D. J., Moyroud, E., Wanke, A., Bonfanti, A., Gatti, S., Summers, A., Burmeister, E., Grube, K., Alexa, A., Kuksa, N., Gardiner, L., Balcerowicz, M., Salmon, J., Yates, B., Riglet, L., Salvi, E., Szovenyi, P., Cheng, S., Tanizawa, Y., Aziz, S., Leebens-Mack, J. H., Schmutz, J., Webber, J., Grimwood, J., Jacquet, C., Dunand, C., Nelson, J. M., Roux, F., Philippe, H., Schornack, S., Bonhomme, M. & Delaux, P.-M. The Marchantia polymorpha pangenome reveals ancient mechanisms of plant adaptation to the environment. *Nat Genet* (2025). doi:10.1038/s41588-024-02071-4

19. Wu, S., Jandrasits, K., Swarts, K., Roetzer, J., Akimcheva, S., Shimamura, M., Hisanaga, T., Berger, F. & Dolan, L. Population genomics of Marchantia polymorpha subsp. ruderalis reveals evidence of climate adaptation. *Current Biology* **35,** 970-980.e3 (2025).

20. Read, D. J., Duckett, J. G., Francis, R., Ligrone, R. & Russell, A. Symbiotic fungal associations in 'lower' land plants. *Phil. Trans. R. Soc. Lond. B* **355,** 815–831 (2000).

21. Read, D. J., Leake, J. R. & Perez-Moreno, J. Mycorrhizal fungi as drivers of ecosystem processes in heathland and boreal forest biomes. *Can. J. Bot.* **82,** 1243–1263 (2004).

22. Bechteler, J., Peñaloza-Bojacá, G., Bell, D., Gordon Burleigh, J., McDaniel, S. F., Christine Davis, E., Sessa, E. B., Bippus, A., Christine Cargill, D., Chantanoarrapint, S., Draper, I., Endara, L., Forrest, L. L., Garilleti, R., Graham, S. W., Huttunen, S., Lazo, J. J., Lara, F., Larraín, J., Lewis, L. R., Long, D. G., Quandt, D., Renzaglia, K., Schäfer-Verwimp, A., Lee, G. E., Sierra, A. M., Von Konrat, M., Zartman, C. E., Pereira, M. R., Goffinet, B. & Villarreal A., J. C. Comprehensive phylogenomic time tree of bryophytes reveals deep relationships and uncovers gene incongruences in the last 500 million years of diversification. *American J of Botany* **110,** e16249 (2023).

23. Harris, B. J., Clark, J. W., Schrempf, D., Szöllősi, G. J., Donoghue, P. C. J., Hetherington, A. M. & Williams, T. A. Divergent evolutionary trajectories of bryophytes and tracheophytes from a complex common ancestor of land plants. *Nat Ecol Evol* **6,** 1634–1643 (2022).

24. Smith, S. E. & Read, D. J. *Mycorrhizal symbiosis*. (2010). at <https://doi.org/10.1016/B978-0-12-370526-6.X5001-6>

25. Li, X.-R., Sun, J., Albinsky, D., Zarrabian, D., Hull, R., Lee, T., Jarratt-Barnham, E., Chiu, C. H., Jacobsen, A., Soumpourou, E., Albanese, A., Kohlen, W., Luginbuehl, L. H., Guillotin, B.,

Lawrensen, T., Lin, H., Murray, J., Wallington, E., Harwood, W., Choi, J., Paszkowski, U. & Oldroyd, G. E. D. Nutrient regulation of lipochitooligosaccharide recognition in plants via NSP1 and NSP2. *Nat Commun* **13,** 6421 (2022).

26. Shi, J., Wang, X. & Wang, E. Mycorrhizal Symbiosis in Plant Growth and Stress Adaptation: From Genes to Ecosystems. *Annu. Rev. Plant Biol.* **74,** 569–607 (2023).

27. Hambleton, S. & Sigler, L. Meliniomyces, a new anamorph genus for root-associated fungi with phylogenetic affinities to Rhizoscyphus ericae (≡ Hymenoscyphus ericae), Leotiomycetes. *Studies in Mycology* **53,** 1–27 (2005).

28. Sigler, L. & Gibas, C. F. C. Utility of a cultural method for identification of the ericoid mycobiont Oidiodendron maius confirmed by ITS sequence analysis. *Studies in Mycology* **53,** 63–74 (2005).

29. Fehrer, J., Réblová, M., Bambasová, V. & Vohník, M. The root-symbiotic *Rhizoscyphus ericae* aggregate and *Hyaloscypha* ( *Leotiomycetes* ) are congeneric: Phylogenetic and experimental evidence. *Studies in Mycology* **92,** 195–225 (2019).

30. Vohník, M., Figura, T. & Réblová, M. Hyaloscypha gabretae and Hyaloscypha gryndleri spp. nov. (Hyaloscyphaceae, Helotiales), two new mycobionts colonizing conifer, ericaceous and orchid roots. *Mycorrhiza* **32,** 105–122 (2022).

31. Paries, M. & Gutjahr, C. The good, the bad, and the phosphate: regulation of beneficial and detrimental plant–microbe interactions by the plant phosphate status. *New Phytologist* **239,** 29–46 (2023).

32. Zhao, B., Jia, X., Yu, N., Murray, J. D., Yi, K. & Wang, E. Microbe-dependent and independent nitrogen and phosphate acquisition and regulation in plants. *New Phytol* **242,** 1507–22 (2024).

33. Kiers, E. T., Duhamel, M., Beesetty, Y., Mensah, J. A., Franken, O., Verbruggen, E., Fellbaum, C. R., Kowalchuk, G. A., Hart, M. M., Bago, A., Palmer, T. M., West, S. A., Vandenkoornhuyse, P., Jansa, J. & Bücking, H. Reciprocal Rewards Stabilize Cooperation in the Mycorrhizal Symbiosis. *Science* **333,** 880–882 (2011).

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

35. Vernié, T., Rich, M., Pellen, T., Teyssier, E., Garrigues, V., Chauderon, L., Medioni, L., Van Beveren, F., Libourel, C., Keller, J., Girou, C., Lefort, C., Le Ru, A., Martinez, Y., Reinhardt, D., Kodama, K., Shimazaki, S., Morel, P., Kyozuka, J., Mbengue, M., Vandenbussche, M. & Delaux, P.-M. Conservation of symbiotic signaling since the most recent common ancestor of land plants. *Proc. Natl. Acad. Sci. U.S.A.* **122,** e2408539121 (2025).

36. Wang, B., Yeun, L. H., Xue, J., Liu, Y., Ané, J. & Qiu, Y. 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 Phytologist* **186,** 514–525 (2010).

37. Bravo, A., York, T., Pumplin, N., Mueller, L. A. & Harrison, M. J. Genes conserved for arbuscular mycorrhizal symbiosis identified through phylogenomics. *Nature Plants* **2,** 15208 (2016).

38. Yano, K., Yoshida, S., Müller, J., Singh, S., Banba, M., Vickers, K., Markmann, K., White, C., Schuller, B., Sato, S., Asamizu, E., Tabata, S., Murooka, Y., Perry, J., Wang, T. L., Kawaguchi, M., Imaizumi-Anraku, H., Hayashi, M. & Parniske, M. CYCLOPS, a mediator of symbiotic intracellular accommodation. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 20540–20545 (2008).

39. 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).

40. Gleason, C., Chaudhuri, S., Yang, T., Muñoz, A., Poovaiah, B. W. & Oldroyd, G. E. D. Nodulation independent of rhizobia induced by a calcium-activated kinase lacking autoinhibition. *Nature* **441,** 1149–1152 (2006).

41. Luginbuehl, L. H. & Oldroyd, G. E. D. Understanding the Arbuscule at the Heart of Endomycorrhizal Symbioses in Plants. *Current Biology* **27,** R952–R963 (2017).

42. Schiessl, K., Lilley, J. L. S., Lee, T., Tamvakis, I., Kohlen, W., Bailey, P. C., Thomas, A., Luptak, J., Ramakrishnan, K., Carpenter, M. D., Mysore, K. S., Wen, J., Ahnert, S., Grieneisen, V. A. & Oldroyd, G. E. D. NODULE INCEPTION Recruits the Lateral Root Developmental Program for Symbiotic Nodule Organogenesis in Medicago truncatula. *Current Biology* **29,** 3657-3668.e5 (2019).

43. Li, F.-W., Nishiyama, T., Waller, M., Frangedakis, E., Keller, J., Li, Z., Fernandez-Pozo, N., Barker, M. S., Bennett, T., Blázquez, M. A., Cheng, S., Cuming, A. C., De Vries, J., De Vries, S., Delaux, P.-M., Diop, I. S., Harrison, C. J., Hauser, D., Hernández-García, J., Kirbis, A., Meeks, J. C., Monte, I., Mutte, S. K., Neubauer, A., Quandt, D., Robison, T., Shimamura, M., Rensing, S. A., Villarreal, J. C., Weijers, D., Wicke, S., Wong, G. K.-S., Sakakibara, K. & Szövényi, P. Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. *Nat. Plants* **6,** 259–272 (2020).

44. Lanfranco, L., Fiorilli, V. & Gutjahr, C. Partner communication and role of nutrients in the arbuscular mycorrhizal symbiosis. *New Phytologist* **220,** 1031–1046 (2018).

45. Wanke, A., van Boerdonk, S., Mahdi, L. K., Wawra, S., Neidert, M., Chandrasekar, B., Saake, P., Saur, I. M. L., Derbyshire, P., Holton, N., Menke, F. L. H., Brands, M., Pauly, M., Acosta, I. F., Zipfel, C. & Zuccaro, A. A GH81-type β-glucan-binding protein enhances colonization by mutualistic fungi in barley. *Curr Biol* **33,** 5071–84 (2023).

46. Breuillin-Sessoms, F., Floss, D. S., Gomez, S. K., Pumplin, N., Ding, Y., Levesque-Tremblay, V., Noar, R. D., Daniels, D. A., Bravo, A., Eaglesham, J. B., Benedito, V. A., Udvardi, M. K. & Harrison, M. J. Suppression of Arbuscule Degeneration in *Medicago truncatula phosphate transporter4* Mutants Is Dependent on the Ammonium Transporter 2 Family Protein AMT2;3. *The Plant Cell* **27,** 1352–1366 (2015).

47. Kawaharada, Y., Kelly, S., Nielsen, M. W., Hjuler, C. T., Gysel, K., Muszyński, A., Carlson, R. W., Thygesen, M. B., Sandal, N., Asmussen, M. H., Vinther, M., Andersen, S. U., Krusell, L., Thirup, S., Jensen, K. J., Ronson, C. W., Blaise, M., Radutoiu, S. & Stougaard, J. Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature* **523,** 308–312 (2015).

48. Kelly, S., Hansen, S. B., Rübsam, H., Saake, P., Pedersen, E. B., Gysel, K., Madland, E., Wu, S., Wawra, S., Reid, D., Sullivan, J. T., Blahovska, Z., Vinther, M., Muszynski, A., Azadi, P.,

Thygesen, M. B., Aachmann, F. L., Ronson, C. W., Zuccaro, A., Andersen, K. R., Radutoiu, S. & Stougaard, J. A glycan receptor kinase facilitates intracellular accommodation of arbuscular mycorrhiza and symbiotic rhizobia in the legume Lotus japonicus. *PLoS Biol* **21,** e3002127 (2023).

49. Liu, C.-W., Breakspear, A., Stacey, N., Findlay, K., Nakashima, J., Ramakrishnan, K., Liu, M., Xie, F., Endre, G., De Carvalho-Niebel, F., Oldroyd, G. E. D., Udvardi, M. K., Fournier, J. & Murray, J. D. A protein complex required for polar growth of rhizobial infection threads. *Nat Commun* **10,** 2848 (2019).

50. Feddermann, N., Duvvuru Muni, R. R., Zeier, T., Stuurman, J., Ercolin, F., Schorderet, M. & Reinhardt, D. The PAM1 gene of petunia, required for intracellular accommodation and morphogenesis of arbuscular mycorrhizal fungi, encodes a homologue of VAPYRIN: Intracellular accommodation of AM fungi. *The Plant Journal* **64,** 470–481 (2010).

51. Pumplin, N., Mondo, S. J., Topp, S., Starker, C. G., Gantt, J. S. & Harrison, M. J. *Medicago truncatula* Vapyrin is a novel protein required for arbuscular mycorrhizal symbiosis. *The Plant Journal* **61,** 482–494 (2010).

52. Murray, J. D., Muni, R. R. D., Torres-Jerez, I., Tang, Y., Allen, S., Andriankaja, M., Li, G., Laxmi, A., Cheng, X., Wen, J., Vaughan, D., Schultze, M., Sun, J., Charpentier, M., Oldroyd, G., Tadege, M., Ratet, P., Mysore, K. S., Chen, R. & Udvardi, M. K. *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*. *The Plant Journal* **65,** 244–252 (2011).

53. Liu, M., Jia, N., Li, X., Liu, R., Xie, Q., Murray, J. D., Downie, J. A. & Xie, F. CERBERUS is critical for stabilization of VAPYRIN during rhizobial infection in *Lotus japonicus*. *New Phytologist* **229,** 1684–1700 (2021).

54. 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 and Cell Physiology* **54,** 1711–1723 (2013).

55. Arrighi, J.-F., Godfroy, O., De Billy, F., Saurat, O., Jauneau, A. & Gough, C. The *RPG* gene of *Medicago truncatula* controls *Rhizobium* -directed polar growth during infection. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 9817–9822 (2008).

56. Griesmann, M., Chang, Y., Liu, X., Song, Y., Haberer, G., Crook, M. B., Billault-Penneteau, B., Lauressergues, D., Keller, J., Imanishi, L., Roswanjaya, Y. P., Kohlen, W., Pujic, P., Battenberg, K., Alloisio, N., Liang, Y., Hilhorst, H., Salgado, M. G., Hocher, V., Gherbi, H., Svistoonoff, S., Doyle, J. J., He, S., Xu, Y., Xu, S., Qu, J., Gao, Q., Fang, X., Fu, Y., Normand, P., Berry, A. M., Wall, L. G., Ané, J.-M., Pawlowski, K., Xu, X., Yang, H., Spannagl, M., Mayer, K. F. X., Wong, G. K.-S., Parniske, M., Delaux, P.-M. & Cheng, S. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* **361,** eaat1743 (2018).

57. Zhang, X., Pumplin, N., Ivanov, S. & Harrison, M. J. EXO70I Is Required for Development of a Sub-domain of the Periarbuscular Membrane during Arbuscular Mycorrhizal Symbiosis. *Current Biology* **25,** 2189–2195 (2015).

58. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18,** 170–175 (2021).

59. Uliano-Silva, M. MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio High Fidelity reads. (2022).

60. 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).

61. Baril, T., Galbraith, J. & Hayward, A. Earl Grey: A Fully Automated User-Friendly Transposable Element Annotation and Analysis Pipeline. *Molecular Biology and Evolution* **41,** msae068 (2024).

62. Gabriel, L., Brůna, T., Hoff, K. J., Ebel, M., Lomsadze, A., Borodovsky, M. & Stanke, M. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res.* **34,** 769–777 (2024).

63. Kim, D. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* **37,** (2019).

64. Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M. & Li, H. Twelve years of SAMtools and BCFtools. *GigaScience* **10,** giab008 (2021).

65. Li, H. Protein-to-genome alignment with miniprot. *Bioinformatics* **39,** btad014 (2023).

66. Brůna, T., Lomsadze, A. & Borodovsky, M. GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. *Genome Res.* **34,** 757–768 (2024).

67. Iwata, H. & Gotoh, O. Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. *Nucleic Acids Research* **40,** e161–e161 (2012).

68. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12,** 59–60 (2015).

69. Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L. & Pertea, M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol* **20,** 278 (2019).

70. Pertea, G. M. & Pertea, M. GFF Utilities: GffRead and GffCompare. (2020). at <10.12688/f1000research.23297.2>

71. Stanke, M., Schöffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7,** 62 (2006).

72. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* **24,** 637–644 (2008).

73. Gabriel, L., Hoff, K. J., Brůna, T., Borodovsky, M. & Stanke, M. TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics* **22,** 566 (2021).

74. Huang, N. & Li, H. compleasm: a faster and more accurate reimplementation of BUSCO. *Bioinformatics* **39,** btad595 (2023).

75. Dainat, J. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format. (2022). at <github.com/NBISweden/AGAT>

76. Blum, M., Andreeva, A., Florentino, L. C., Chuguransky, S. R., Grego, T., Hobbs, E., Pinto, B. L., Orr, A., Paysan-Lafosse, T., Ponamareva, I., Salazar, G. A., Bordin, N., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunic, I., Llinares-López, F., Marchler-Bauer, A., Meng-Papaxanthos, L., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Piovesan, D., Rivoire, C., Sigrist, C. J. A., Thanki, N., Thibaud-Nissen, F., Thomas, P. D., Tosatto, S. C. E., Wu, C. H. & Bateman, A. InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Research* **53,** D444–D456 (2025).

77. Lee, E., Helt, G. A., Reese, J. T., Munoz-Torres, M. C., Childers, C. P., Buels, R. M., Stein, L., Holmes, I. H., Elsik, C. G. & Lewis, S. E. Web Apollo: a web-based genomic annotation editing platform. *Genome Biol* **14,** R93 (2013).

78. Chen, H. & Zwaenepoel, A. in *Polyploidy* (ed. Van De Peer, Y.) **2545,** 3–18 (Springer US, 2023).

79. Tang, H., Krishnakumar, V., Zeng, X., Xu, Z., Taranto, A., Lomas, J. S., Zhang, Y., Huang, Y., Wang, Y., Yim, W. C., Zhang, J. & Zhang, X. JCVI: A versatile toolkit for comparative genomics analysis. *iMeta* **3,** e211 (2024).

80. Smith, G. S., Johnston, C. M. & Cornforth, I. S. COMPARISON OF NUTRIENT SOLUTIONS FOR GROWTH OF PLANTS IN SAND CULTURE. *New Phytologist* **94,** 537–548 (1983).

81. Atlas, M. R. *Handbook of Microbiological Media*. (CRC Press, 2010).

82. Abràmoff, D. M. D. Image Processing with ImageJ. *Biophotonics International* **11,** 36–42 (2004).

83. Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E. & Notredame, C. Nextflow enables reproducible computational workflows. *Nat Biotechnol* **35,** 316–319 (2017).

84. Ewels, P. A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M. U., Di Tommaso, P. & Nahnsen, S. The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol* **38,** 276–278 (2020).

85. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28,** 3211–3217 (2012).

86. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).

87. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15,** 550 (2014).

88. Sayols, S., Scherzinger, D. & Klein, H. dupRadar: a Bioconductor package for the assessment of PCR artifacts in RNA-Seq data. *BMC Bioinformatics* **17,** 428 (2016).

89. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. (2011).

90. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Batut, P., Chaisson, M. & Gingeras, T. R. STAR: ultrafast universal RNA-seq aligner. (2012).

91. Okonechnikov, K., Conesa, A. & García-Alcalde, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32,** 292–294 (2016).

92. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28,** 2184–2185 (2012).

93. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14,** 417–419 (2017).

94. Morgan, M., Obenchain, V., Hester, J. & Pagès, H. SummarizedExperiment: A container (S4 class) for matrix-like assays. (2024). at <bioconductor.org/packages/SummarizedExperiment>

95. Love, M. I., Soneson, C., Hickey, P. F., Johnson, L. K., Pierce, N. T., Shepherd, L., Morgan, M. & Patro, R. Tximeta: Reference sequence checksums for provenance identification in RNA-seq. *PLoS Comput Biol* **16,** e1007664 (2020).

96. Chen, Y., Chen, L., Lun, A. T. L. & Baldoni, P. L. edgeR 4.0: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. (2024).

97. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X. & Yu, G. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2,** 100141 (2021).

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

99. Emms, D. M. & Kelly, S. STAG: Species Tree Inference from All Genes. Preprint at https://doi.org/10.1101/267914 (2018)

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

101. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30,** 772–780 (2013).

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

103. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **5,** e9490 (2010).

104. Edgar, R. C. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat Commun* **13,** 6968 (2022).

105. Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A. & Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37,** 1530–1534 (2020).

106. Yang, Z., Nielsen, R. & Hasegawa, M. Models of amino acid substitution and applications to mitochondrial protein evolution. *Molecular Biology and Evolution* **15,** 1600–1611 (1998).

107. Yang, Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution* **15,** 568–573 (1998).

108. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* **34,** W609–W612 (2006).

109. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* **24,** 1586–1591 (2007).

110. Mai, U. & Mirarab, S. TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics* **19,** 272 (2018).

111. Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., Du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M. A., Wu, C.-H., Xie, D., Zhang, C., Stadler, T. & Drummond, A. J. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* **15,** e1006650 (2019).

112. Douglas, J., Zhang, R. & Bouckaert, R. Adaptive dating and fast proposals: Revisiting the phylogenetic relaxed clock model. *PLoS Comput Biol* **17,** e1008322 (2021).

113. Hu, R., Li, X., Hu, Y., Zhang, R., Lv, Q., Zhang, M., Sheng, X., Zhao, F., Chen, Z., Ding, Y., Yuan, H., Wu, X., Xing, S., Yan, X., Bao, F., Wan, P., Xiao, L., Wang, X., Xiao, W., Decker, E. L., Van Gessel, N., Renault, H., Wiedemann, G., Horst, N. A., Haas, F. B., Wilhelmsson, P. K. I., Ullrich, K. K., Neumann, E., Lv, B., Liang, C., Du, H., Lu, H., Gao, Q., Cheng, Z., You, H., Xin, P., Chu, J., Huang, C.-H., Liu, Y., Dong, S., Zhang, L., Chen, F., Deng, L., Duan, F., Zhao, W., Li, K., Li, Z., Li, X., Cui, H., Zhang, Y. E., Ma, C., Zhu, R., Jia, Y., Wang, M., Hasebe, M., Fu, J., Goffinet, B., Ma, H., Rensing, S. A., Reski, R. & He, Y. Adaptive evolution of the enigmatic Takakia now facing climate change in Tibet. *Cell* **186,** 3558-3576.e17 (2023).

114. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology* **67,** 901–904 (2018).

115. Engler, C. & Marillonnet, S. in *DNA Cloning and Assembly Methods* (eds. Valla, S. & Lale, R.) **1116,** 119–131 (Humana Press, 2014).

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

117. Andriankaja, A., Boisson-Dernier, A., Frances, L., Sauviac, L., Jauneau, A., Barker, D. G. & De Carvalho-Niebel, F. AP2-ERF Transcription Factors Mediate Nod Factor–Dependent Mt *ENOD11* Activation in Root Hairs via a Novel *cis* -Regulatory Motif. *The Plant Cell* **19,** 2866–2885 (2007).

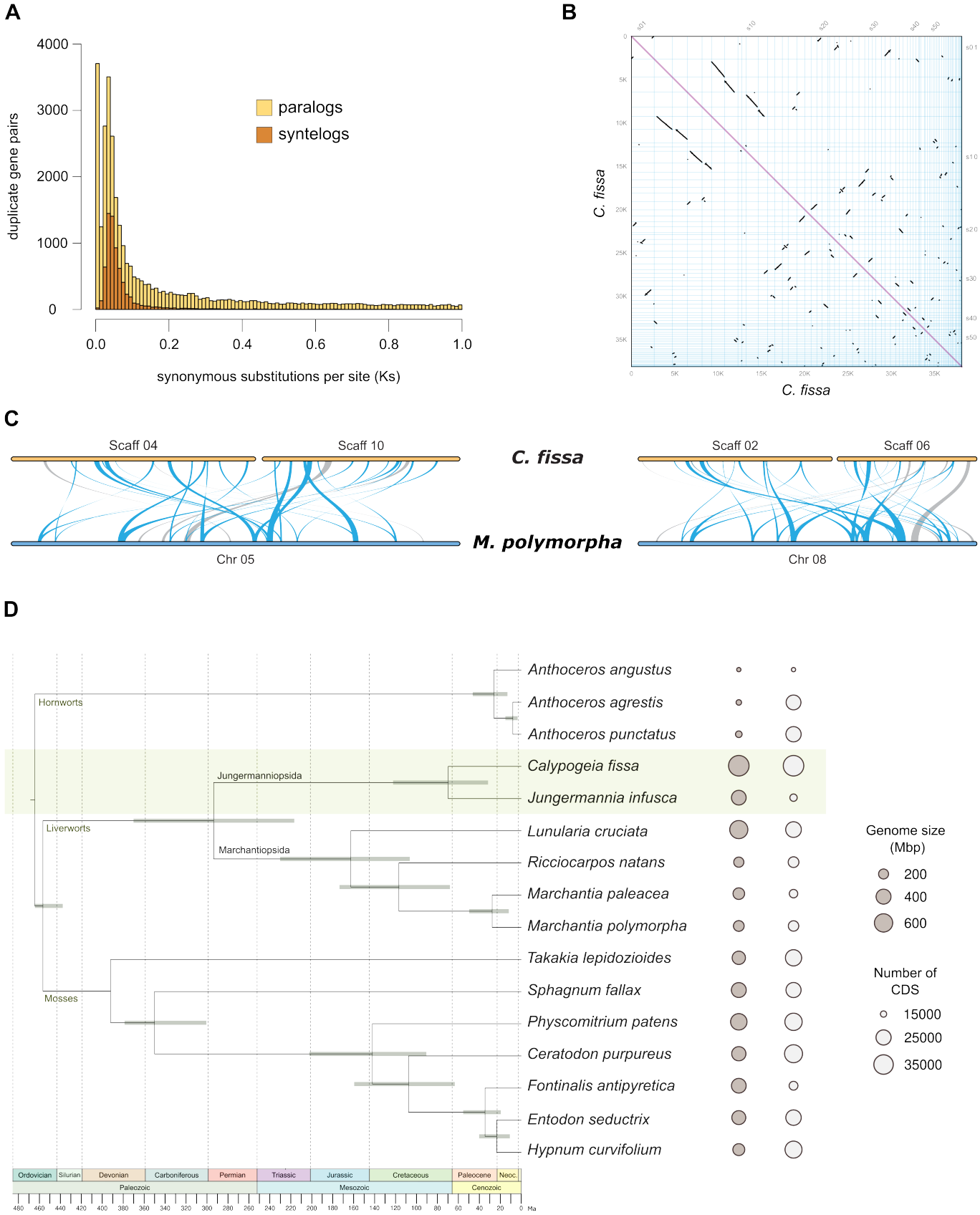



**Figure 1. A recent whole-genome duplication in the leafy liverwort *Calypogeia fissa*.**
**(A)** Distribution of synonymous substitutions per site (*Ks*) between paralog pairs and syntelogs in the *C. fissa* genome. **(B)** Intragenomic synteny analysis identifies conserved collinear gene arrangements within the *C. fissa* genome, revealing 137 collinear blocks comprising 7,087 syntelogs and covering 37.2% of the *C. fissa* genome. **(C)** Synteny between *C. fissa* and *Marchantia polymorpha*. **(D)** Time-calibrated phylogeny of bryophyte species. Node bars represent the 95% highest posterior density (HPD) intervals. The root of the tree and the crown node of mosses were fixed to the mean ages (see Methods). Circles on the right are proportional to genome sizes and the number of coding sequences (CDS) in the genomes of the corresponding species.

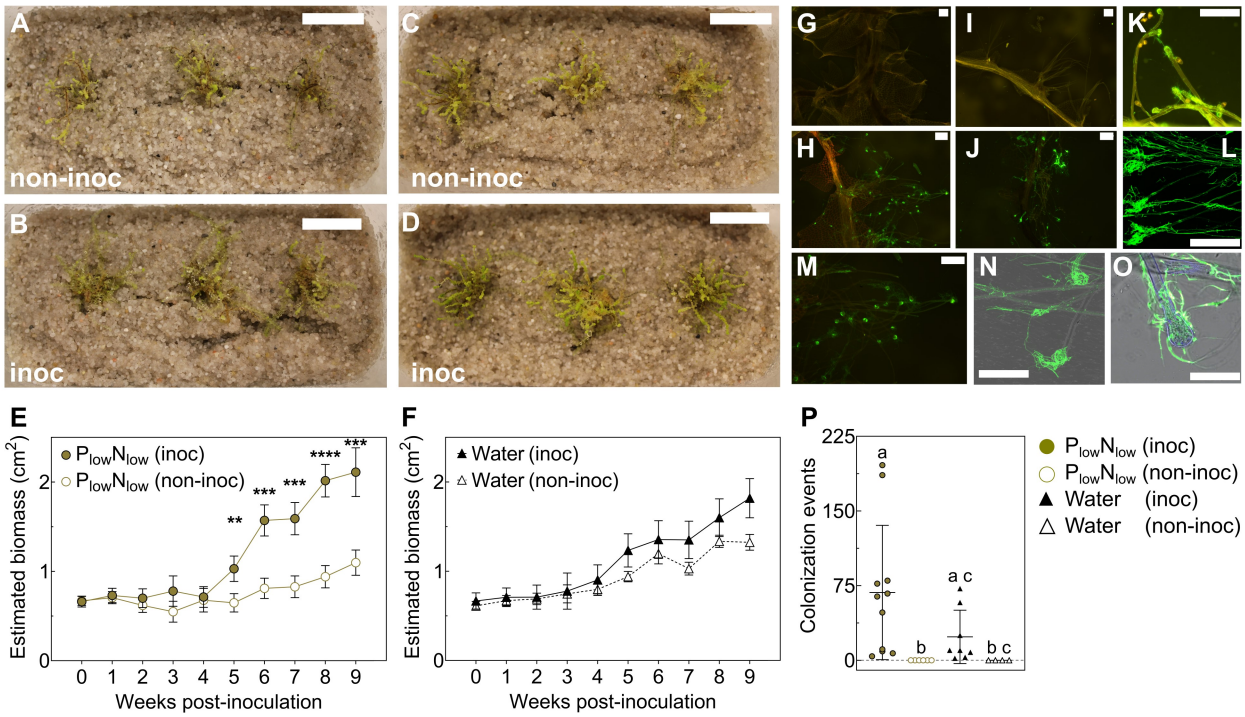



**Figure 2. The establishment of ericoid mycorrhizae in *Calypogeia fissa* is maximized under conditions of both phosphate and nitrogen deficiency. (A-F)** Representative images of *C. fissa* plants growing in mock-inoculated (non-inoc) and inoculated (inoc with *Hyaloscypha hepaticicola*) treatments, as indicated, under phosphate (P) and nitrogen (N) deficiency conditions (A, B), or water (D, E) (see Suppl. Table 1). Scale bars: 2 cm. Estimated growth of *C. fissa* in non-inoc and inoc treatments supplied with a low P/N solution (E) or water only (F). Data points represent the mean ± SEM ($n$ = 27) of technical replicates per time-point from one representative experiment of a total of three independent experiments. **(G-M)** Representative fluorescence microscopy photographs of *C. fissa* stems harvested from non-inco and inoc treatments and stained with Wheat Germ Agglutinin (WGA). Plants were photograph and harvested after 9 weeks of growth. Scale bars: 200 µm. **(N-O)** Confocal microscopy images of samples shown in (M), and further stained with Congo red (blue- fluorescent signal), showing the development of swollen-tip rhizoids where *H. hepaticicola* is intracellularly accommodated (green-fluorescent signal). Scale bars: 200 µm (N), 75 µm (O), and 50 µm (I). **(P)** Total number of swollen and colonized rhizoids per one stem in non-inoc and inoc treatments as indicated in (see also E, F). Data is based on individual quantification of intracellular colonized rhizoids stained with WGA under epifluorescence and confocal microscopes. Data points represent the mean ± SD ($n$ = 4 to 11) of individual technical stem replicates from one representative experiment. In (E, F), different characters indicate statistically significant differences based on a Welch's (double-tailed) $t$-test ($P < 0.05$). See also Supplementary Tables 3.

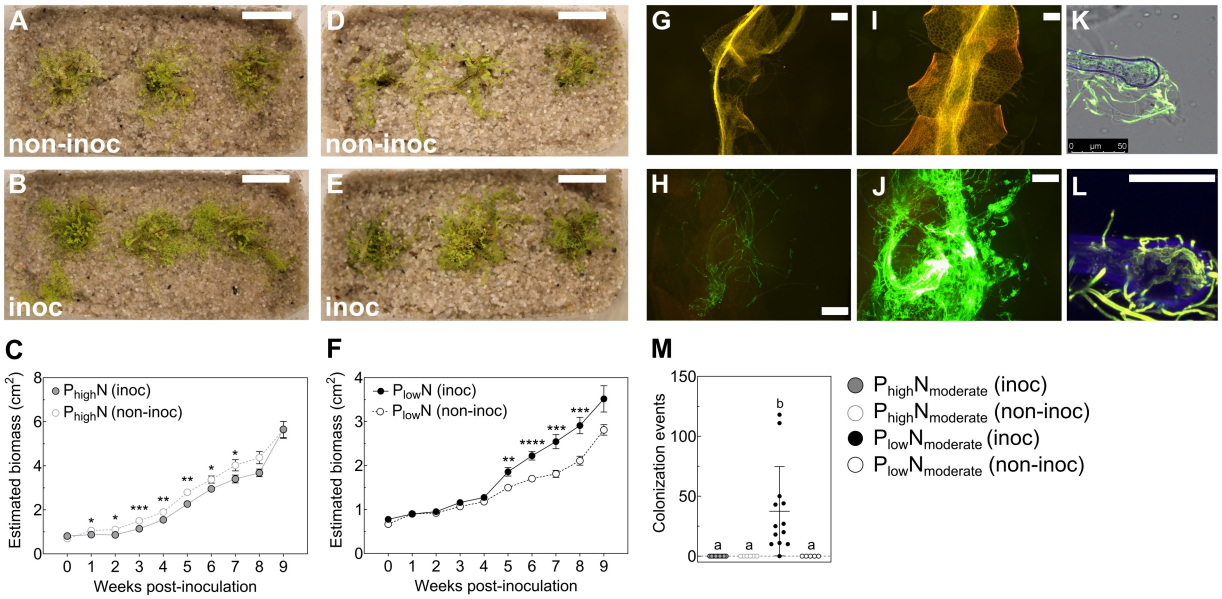



**Figure 3. The establishment of ericoid mycorrhizae in *Calypogeia fissa* is promoted under phosphate limiting conditions and is restricted by high phosphate and moderate N supply.** **(A-F)** Representative images of *C. fissa* plants growing in mock-inoculated (non-inoc) and inoculated (inoc with *Hyaloscypha hepaticicola*) treatments, as indicated, under conditions of high phosphate (P) and moderate nitrogen (N) supply (A, B), or low P and moderate N supply (D, E) (see Supplementary Table 1). Scale bars: 2 cm. Estimated growth of *C. fissa* in non-inoculated and inoculated treatments under the indicated nutrient conditions (C, F). Data points represent the mean ± SEM ($n$ = 27) of technical replicates per time-point from one representative experiment of a total of three independent experiments. **(G-L)** Representative fluorescence microscopy photographs of *C. fissa* stems harvested from mock-inoculated and inoculated treatments and stained with Wheat Germ Agglutinin (WGA, see Methods). Plants were photograph and harvested after 8 weeks of growth (Methods). Scale bars: 200 µm (G, J). Confocal microscopy images of samples shown in (K, L), were further stained with Congo red (blue-fluorescent signal, see also Fig. 2). Scale bar in (L): 25 µm. **(M)** Total number of swollen and colonized rhizoid tips per one stem in non-inoc and inoc treatments as indicated in (C, F).
Data is based on individual quantification of intracellular colonized rhizoids stained with WGA under epifluorescence and confocal microscopes. Data points represent the mean ± SD ($n$ = 4 to 13) of technical stem replicates from one representative experiment. In (C, F), different characters indicate statistically significant differences between inoculated and non-inoculated treatments, based on a Welch's (double-tailed) $t$-test ($P < 0.05$) (C) or based on a Kruskal-Wallis test with post-hoc Dunn's test. See also Supplementary Tables 3.

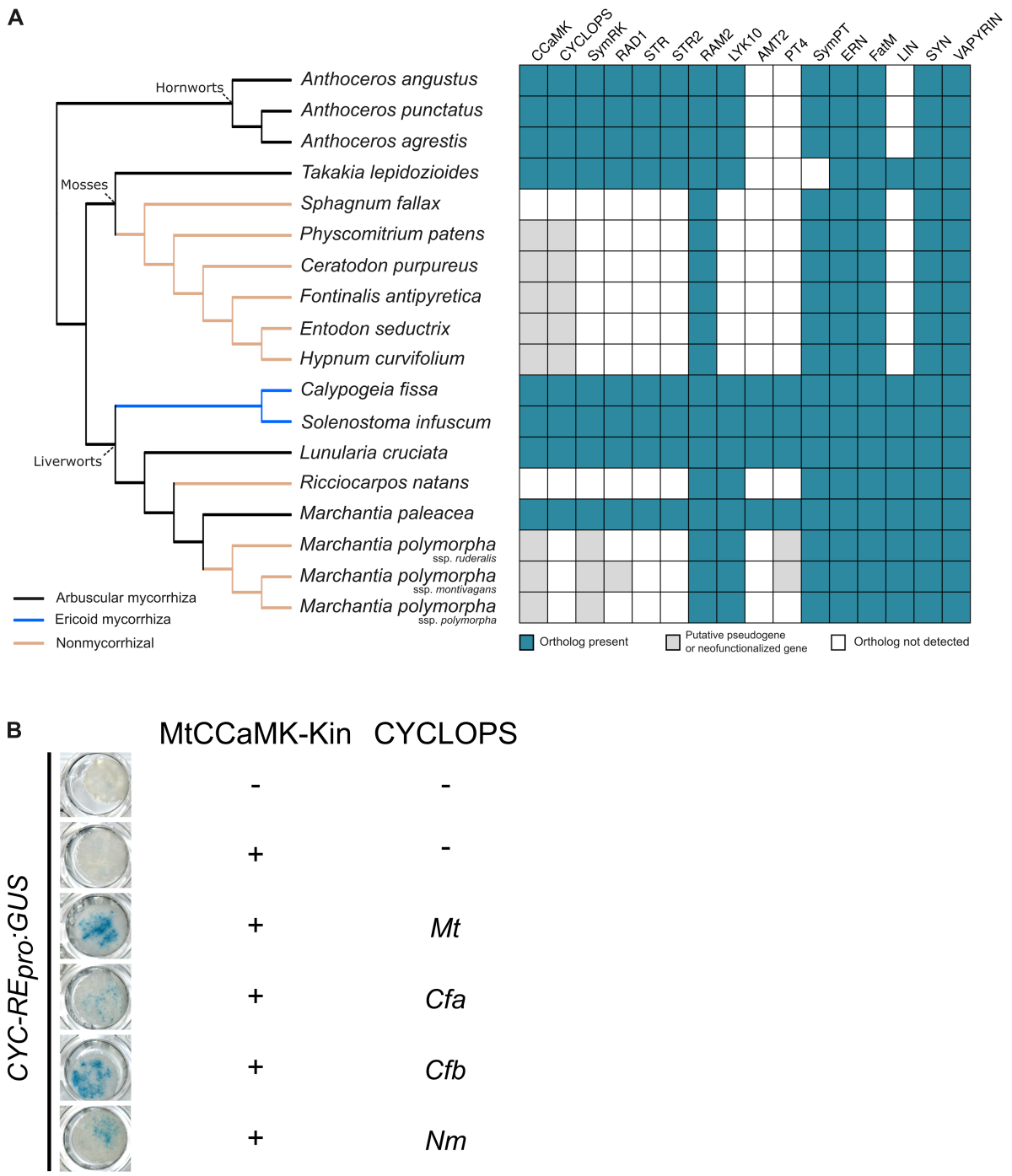



**Figure 4. Conservation and functional validation of ortholog endosymbiotic-related genes in ericoid mycorrhizae-forming leafy liverworts. (A)** Conservation of key genes with known roles during mycorrhizae in the genomes of bryophyte species. **(B)** Conserved transactivation of the *CYC-REpro:GUS* reporter construct by the transcription factor CYCLOPS in Jungermanniales. *Nicotiana benthamiana* leaves were infiltrated with the *CYC-REpro:GUS* construct alone, or with an overexpressing *MtCCaMK* construct in combination with overexpressor constructs of different CYCLOPS orthologs from *Medicago truncatula* (*Mt)*, *Calypogeia fissa* (*Cf*), or *Scapania nemorosa* (*Nm*). Leaves were incubated for 48h with the GUS-staining solution and cleared with ethanol. One representative leaf disc out of six discs is shown for each combination.

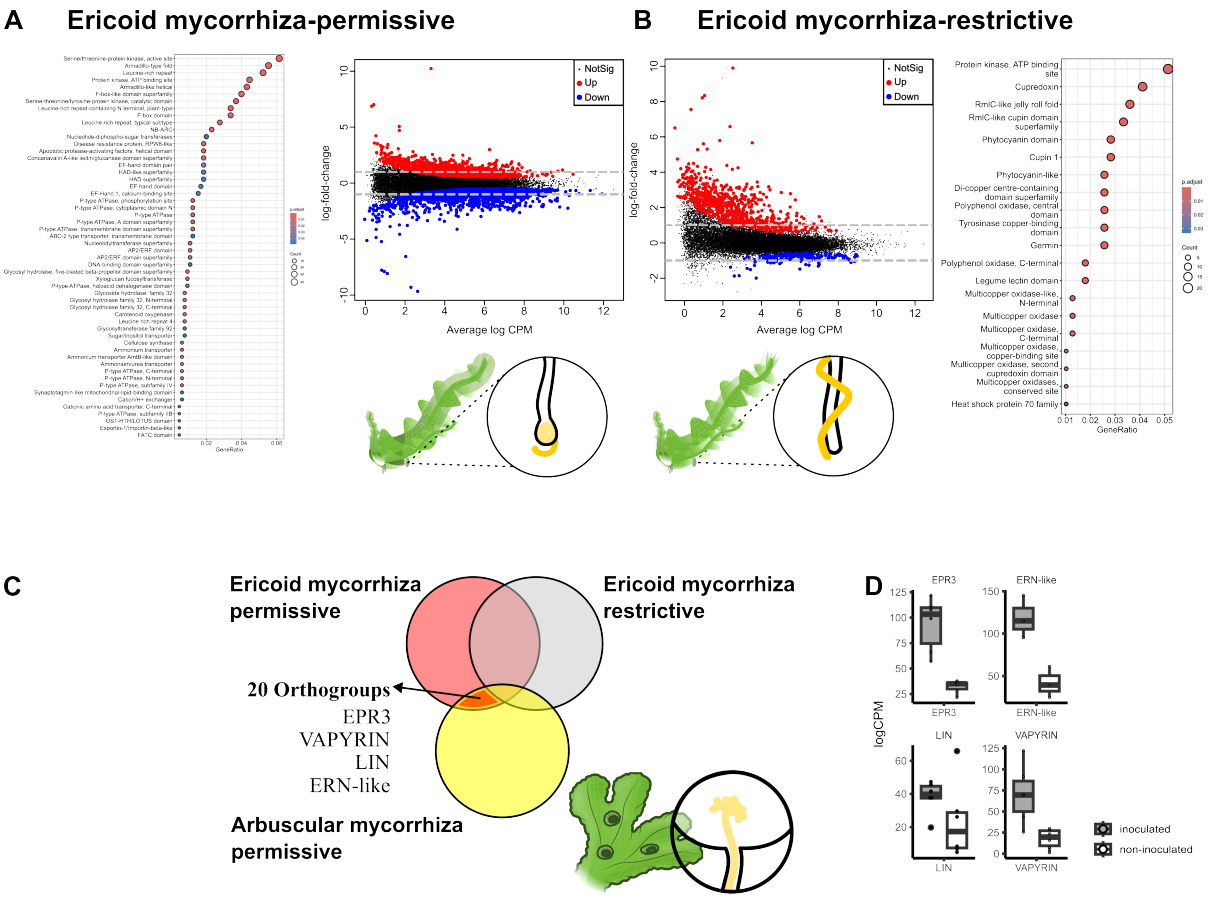
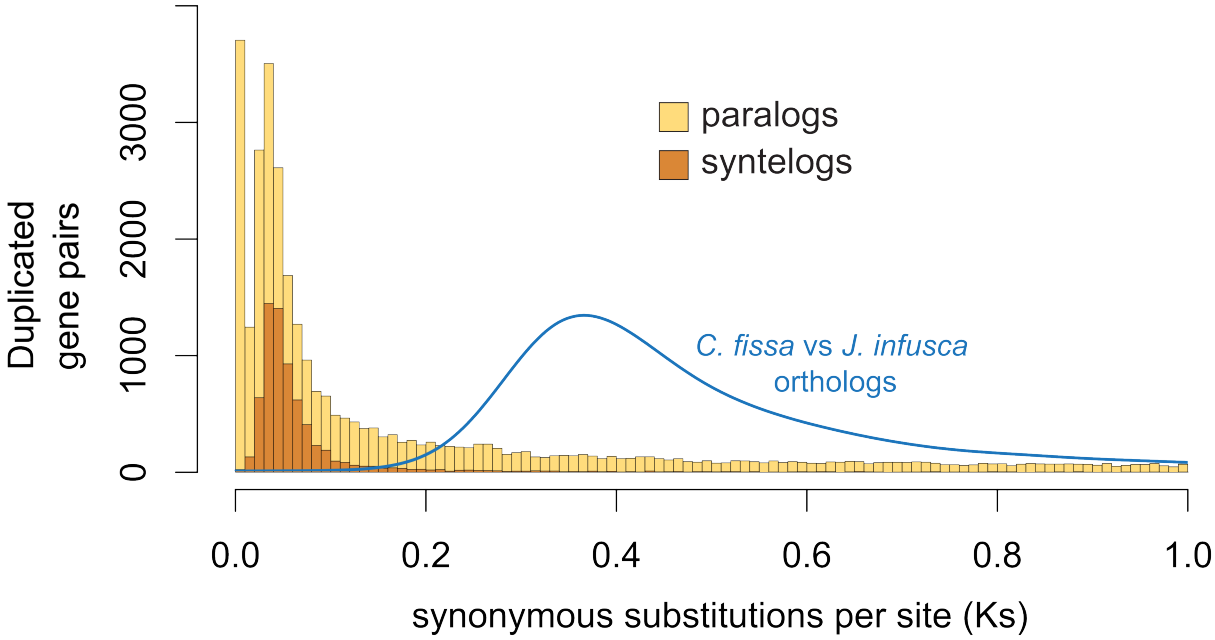
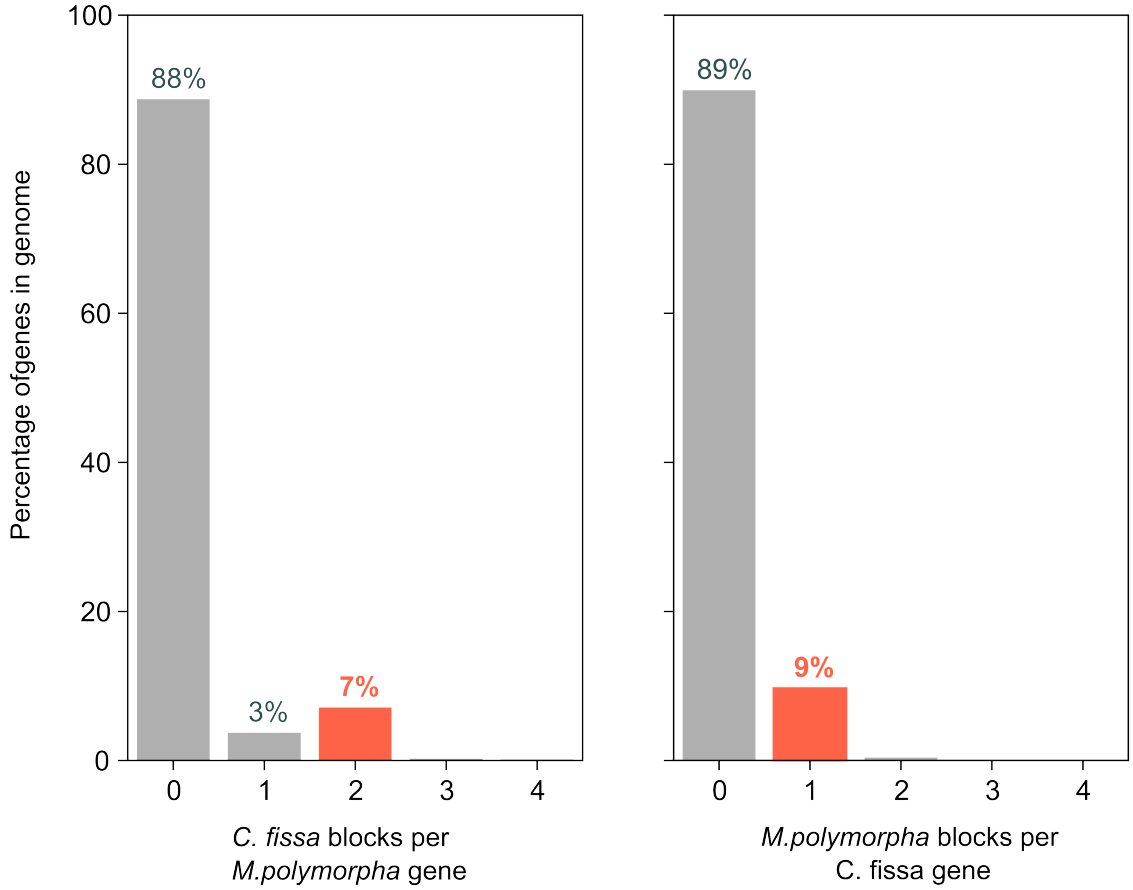


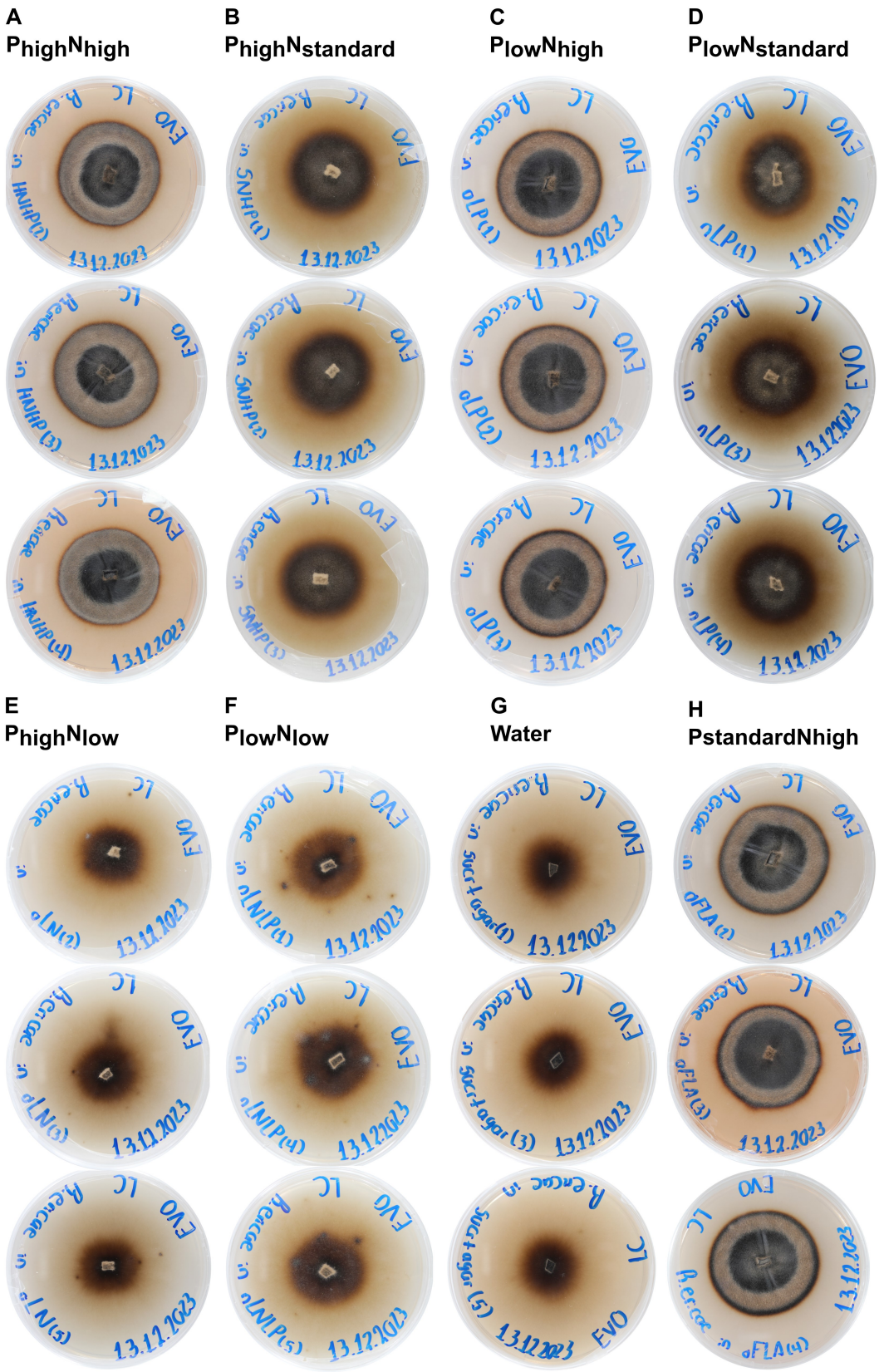

**Figure 5. Transcriptional reprogramming in *Calypogeia fissa* during ericoid mycorrhizae with *Hyaloscypha hepaticicola*. (A)** Differential gene expression analysis comparing mock-treated and *H. hepaticicola*-inoculated *C. fissa* plants under ericoid mycorrhiza-permissive conditions. A total of 787 genes were upregulated, and 743 genes were downregulated in inoculated plants relative to mock-treated controls. **(B)** Differential gene expression analysis under ericoid mycorrhiza-permissive restrictive conditions. In this scenario, 566 genes were upregulated, and 27 genes were downregulated in inoculated plants compared to mock-treated controls. **(C)** Comparative transcriptomic analysis of ericoid mycorrhiza-permissive and restrictive conditions alongside arbuscular mycorrhiza-permissive conditions. The orange-overlapping area in the Venn diagram depicts the conserved expressed gene cluster that was found to be upregulated during both ericoid and arbuscular mycorrhizae. **(D)** Expression bar plots displaying four well-known genes with key functions in endosymbiotic interactions, all of which were found to be upregulated in both ericoid and mycorrhizal symbiosis under permissive conditions. Cross-comparison analysis were done as described in the Methods section.



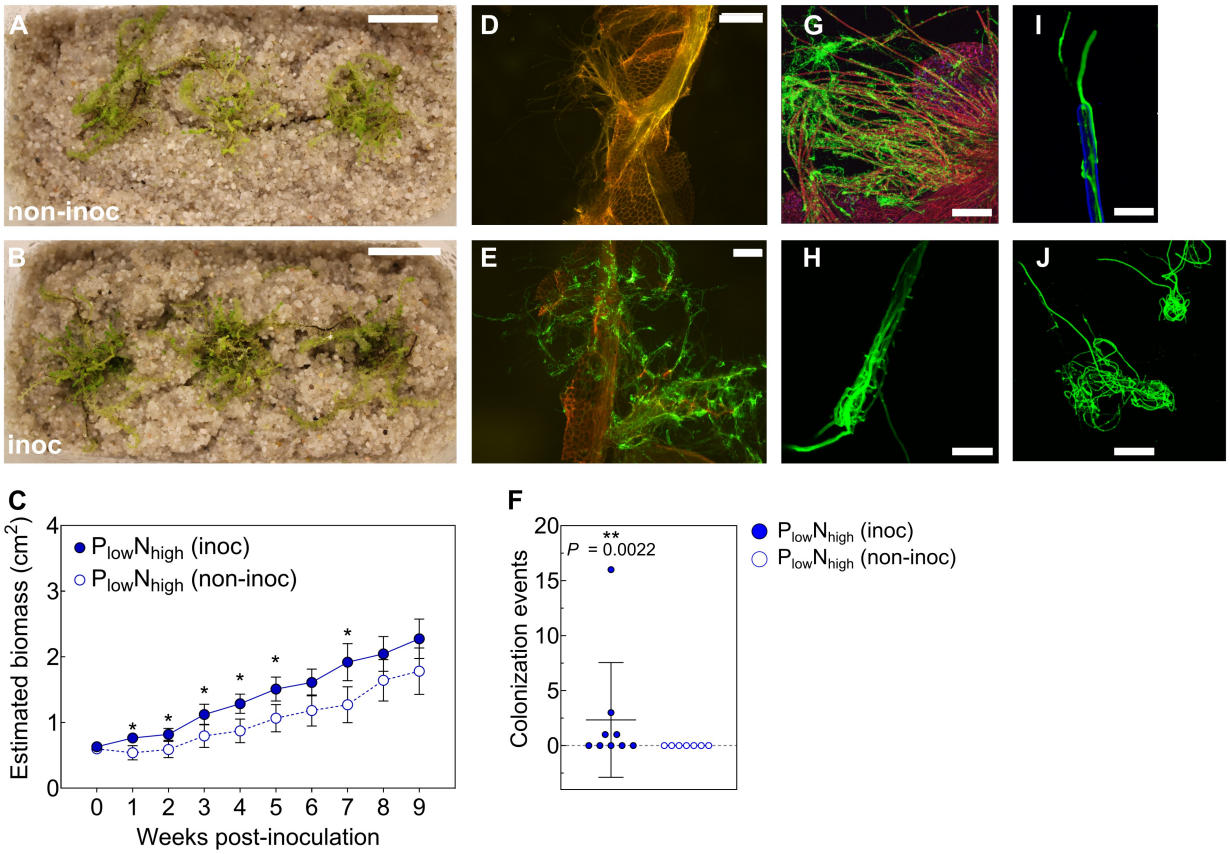

**Supplementary Fig. 1. Orthologous Ks distribution in *Calypogeia fissa* and *Solenostoma infuscum* provides evidence for a recent Whole Genome Duplication (WGD) in *C. fissa*.**
The Ks (synonymous substitution rate) distribution of orthologous genes between *C. fissa* and *S. infuscum* was assessed to investigate the timing of WGD events. The Ks plot shows a prominent peak around Ks = 0.35, indicating a recent WGD event in *C. fissa*. This peak is consistent with a scenario in which WGD occurred after the divergence between *C. fissa* and *S. infuscum*. The distribution suggests that the recent WGD in *C. fissa* predates the split between these two species.



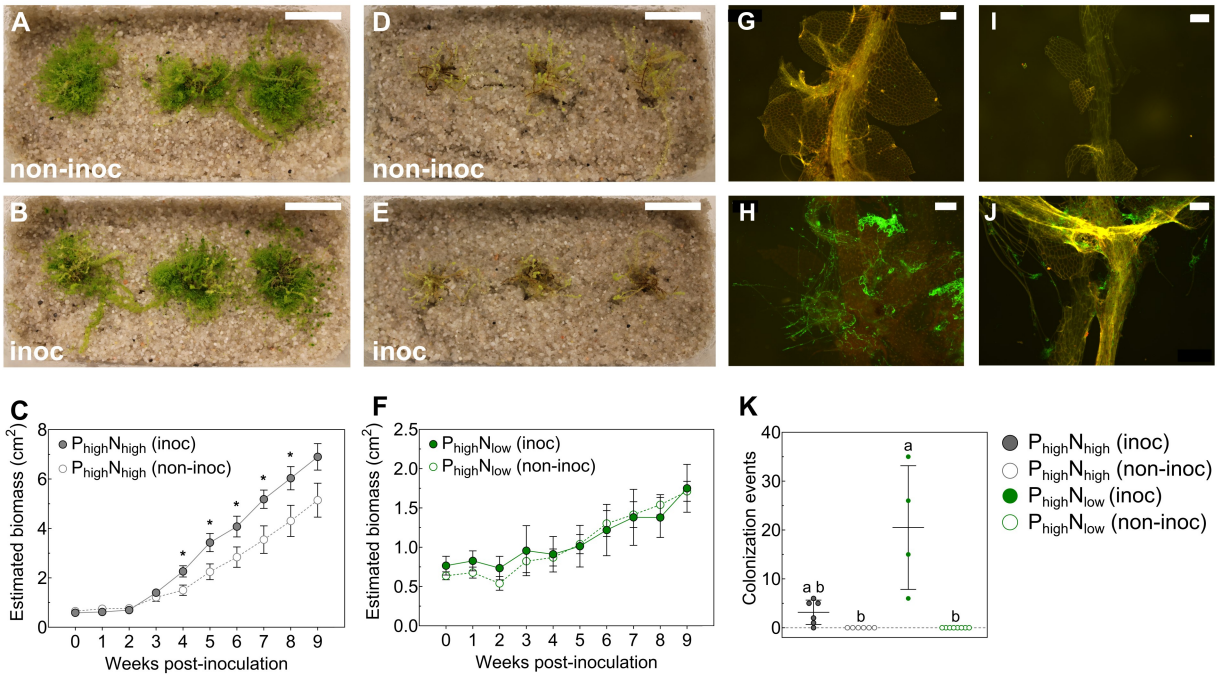

**Supplementary Figure 2. Intergenomic synteny comparison between *Calypogeia fissa* and *Marchantia polymorpha* reveals limited collinearity.** An intergenomic comparison of synteny between *C. fissa* and the distantly related species *M. polymorpha* was conducted to assess the conservation of gene order. The analysis identified relatively few collinear gene pairs, with approximately 10% of genes in each species showing synteny. This limited collinearity is possibly due to the deep phylogenetic divergence between the two species.




**Supplementary Figure 3. Growth of *Hyaloscypha hepaticicola* on agar plates supplemented with variable phosphate and nitrogen nutrient regimes. (A-G)** Representative images of *H. hepaticicola* grown under the indicated nutrient conditions (A-F) used for physiological characterization of *C. fissa* (see Supplementary Table 1) or a basal medium consisting of water, 1.5% (w/v) sucrose, and 0.8% (w/v) Bacto agar (see Methods for details). Plates were incubated at 17°C in darkness and photographed at week 23.

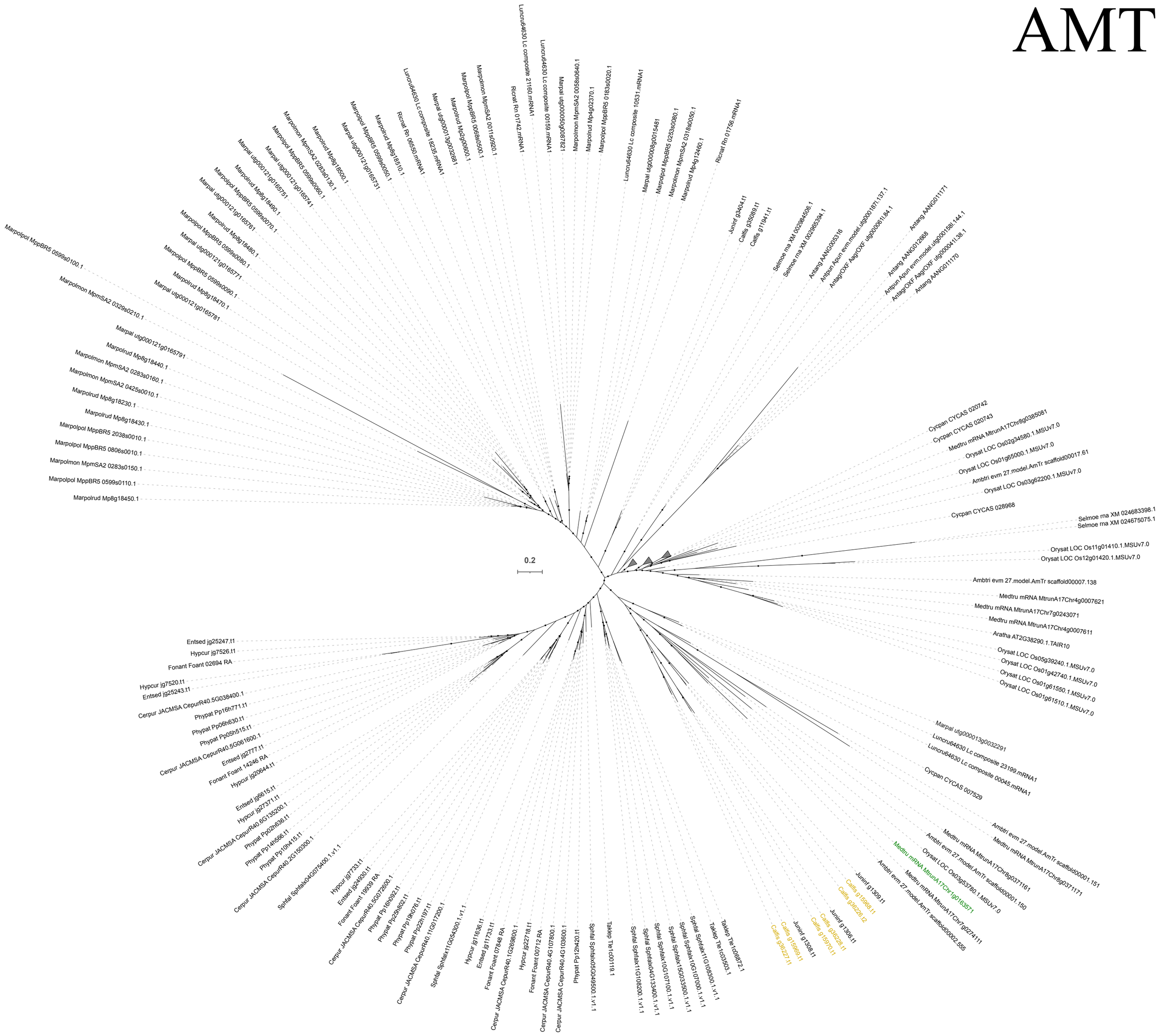



**Supplementary Figure 4. Establishment of ericoid mycorrhizae in *Calypogeia fissa* under phosphate limiting conditions. (A-B)** Representative images of *C. fissa* growing in mock-inoculated (non-inoc) and inoculated treatments as indicated, under conditions of low phosphate (P) and high Nitrogen (N) supply (see Figures 2 and 3). Scale bars: 2 cm. **(C)** Estimated growth of *C. fissa* in non-inoc and inoc treatments, in low P conditions as indicated. Data points represent the mean ± SEM ($n$ = 27) of technical replicates from one representative experiment of a total of three independent experiments. **(D-E)** Representative fluorescence microscopy photographs of *C. fissa* stems harvested from non-inoc and inoc treatments and stained with Wheat Germ Agglutinin (WGA, see Methods). Plants were photograph and harvested after 9 weeks of growth. Scale bars: 200 µm. **(F)** Total number of swollen and colonized rhizoid tips per one stem in non-inoc and inoc treatments as indicated in (C). .Data points represent the mean ± SD ($n$ = 7 to 13) of technical stem replicates from one representative experiment. In (C, F), different characters indicate statistically significant differences based on a Welch's (double-tailed) $t$-test ($P < 0.05$). $P$ values in (C) are based on a Kruskal-Wallis test with a post-hoc Dunn's test. **(G-J)** Confocal microscopy images stained with WGA and Congo red (blue-fluorescent signal), showing intracellular colonization by *H. hepaticicola* (green-fluorescent signal) and the development of swollen rhizoid tips in *C. fissa* stems. Scale bar in (G): 200 µm, (H-J): 25 µm.

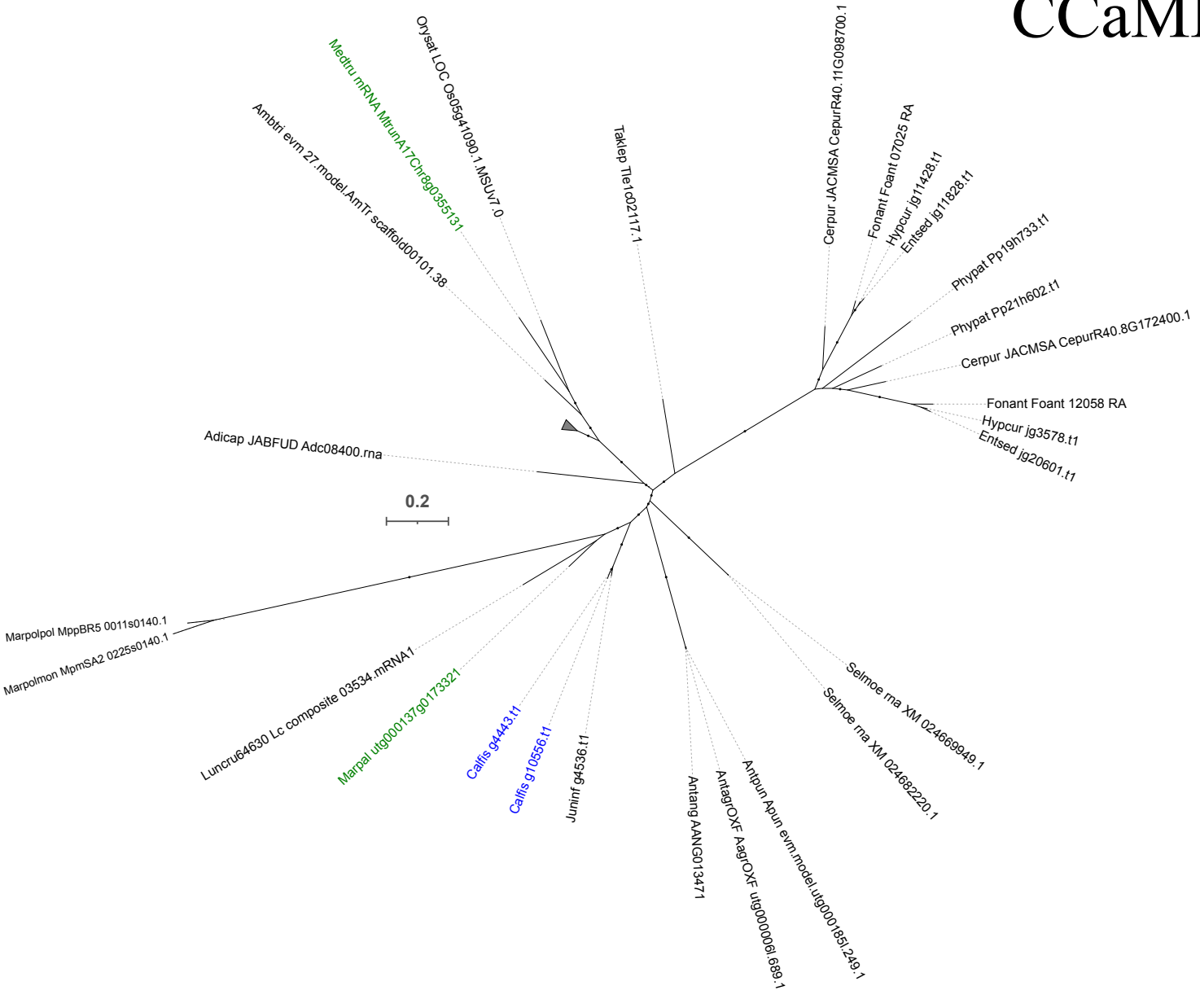



**Supplementary Figure 5. The establishment of ericoid mycorrhizae in *Calypogeia fissa* is restricted under conditions of high phosphate and limited N supply. (A-F)** Representative images of *C. fissa* growing in mock-inoculated (non-inoc) and inoculated (inoc) treatments, as indicated, under conditions of high phosphate (P) and low N supply (D, E) (see also Figs. 2 and 3). Scale bars: 2 cm. Estimated growth of *C. fissa* in non-inoc and inoc treatments under the indicated nutrient conditions (C-F). Data points represent the mean ± SEM ($n$ = 12) of technical replicates from one representative experiment. **(G-L)** Representative fluorescence microscopy photographs of *C. fissa* stems harvested from non-inoculated and inoculated treatments and stained with WGA. Plants were photographed and harvested after 8 weeks of growth (Methods). Scale bars: 200 µm (G-J). **(M)** Total number of swollen and colonized rhizoid tips per one stem in non-inoc and inoc treatments as indicated in (C, F). Data is based on individual quantification of intracellular colonized rhizoids stained with WGA . Data points represent the mean ± SD ($n$ = 4 to 8) of technical stem replicates from one representative experiment. In (C, F), different characters indicate statistically significant differences between inoculated and non-inoculated treatments, based on a Welch's (double-tailed) $t$-test ($P < 0.05$).

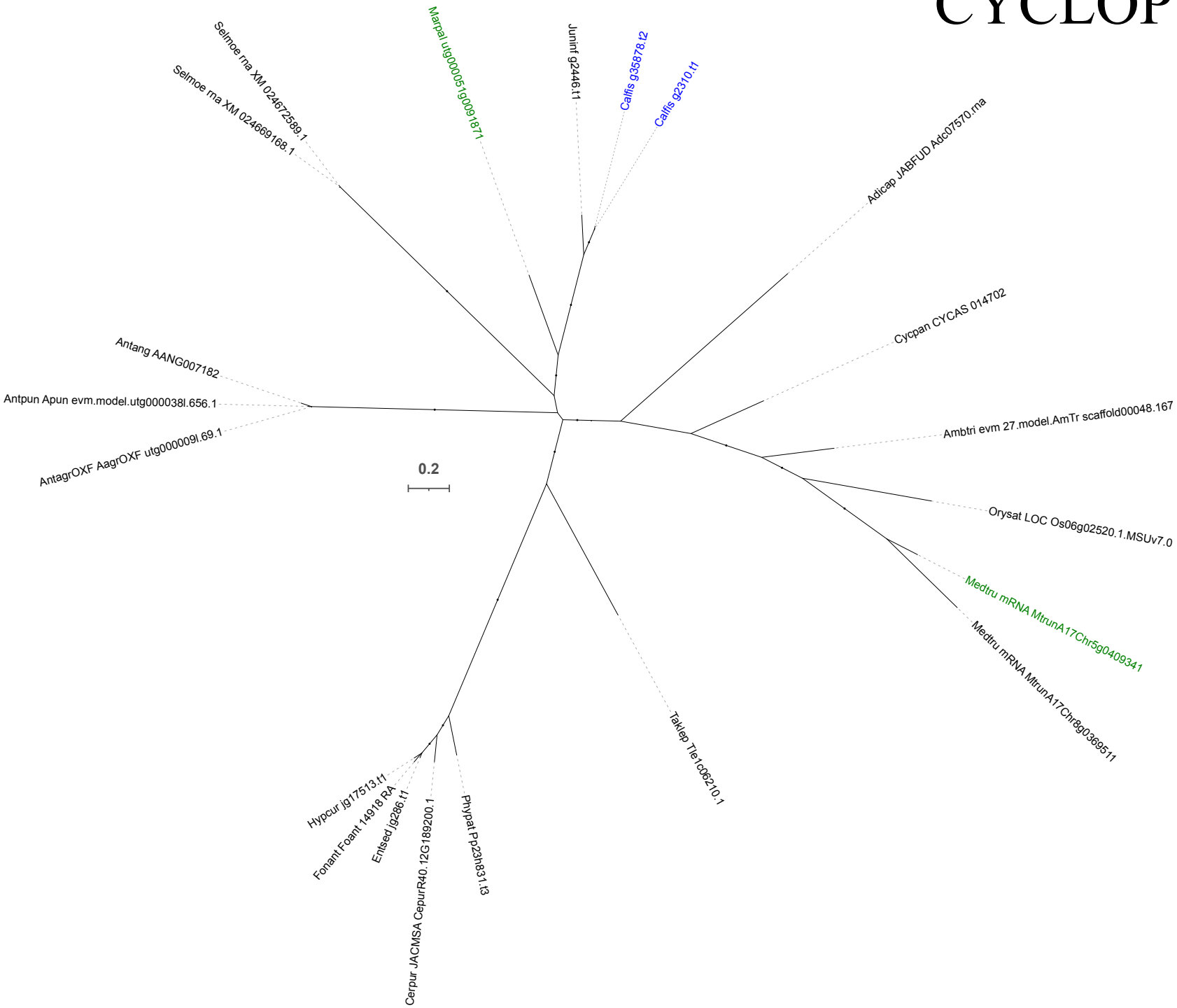
# AMT2


0.2
Luncru64630 Lc composite 21160.mRNA1
Luncru64630 Lc composite 00159.mRNA1
Marpal utg000006g0000001.1
Marpolmon MpmSA2 0058s0640.1
Marpolrud Mp4g02370.1
Marpolpol MppBR5 0183s0020.1
Luncru64630 Lc composite 10531.mRNA1
Marpal utg000008g0015481
Marpolpol MppBR5 0253s0080.1
Marpolmon MpmSA2 0318s0050.1
Marpolrud Mp4g12460.1
Ricnat Rn 01756.mRNA1
Juninf g34041.t1
Calfis g35069.t1
Calfis g11941.t1
Selmoe rna XM 002984506.1
Selmoe rna XM 002965394.1
Antang AANG003316
Antpun Apun evm.model.utg000187l.137.1
AntagrOXF AagrOXF utg000061l.84.1
Antang AANG011171
Antang AANG012668
Antpun Apun evm.model.utg000158l.144.1
AntagrOXF AagrOXF utg000041l.38.1
Antang AANG011170
Ricnat Rn 01742.mRNA1
Marpolmon MpmSA2 0011s0920.1
Marpolpol MppBR3 0068s0500.1
Marpolrud Mp2g00600.1
Marpal utg000013g0032681
Luncru64630 Lc composite 18235.mRNA1
Ricnat Rn 06550.mRNA1
Marpolrud Mp8g18510.1
Marpolpol MppBR5 0599s0050.1
Marpal utg000121g0165731
Marpolrud Mp8g18500.1
Marpolmon MpmSA2 0283s0130.1
Marpolpol MppBR5 0599s0060.1
Marpal utg000121g0165741
Marpal utg000121g0165751
Marpolrud Mp8g18490.1
Marpolpol MppBR5 0599s0070.1
Marpal utg000121g0165761
Marpolrud Mp8g18480.1
Marpolpol MppBR5 0599s0080.1
Marpal utg000121g0165771
Marpolpol MppBR5 0599s0090.1
Marpolrud Mp8g18470.1
Marpal utg000121g0165781
Marpolpol MppBR5 0599s0100.1
Marpolmon MpmSA2 0329s0210.1
Marpal utg000121g0165791
Marpolrud Mp8g18440.1
Marpolmon MpmSA2 0283s0160.1
Marpolmon MpmSA2 0425s0010.1
Marpolrud Mp8g18230.1
Marpolrud Mp8g18430.1
Marpolpol MppBR5 2038s0010.1
Marpolpol MppBR5 0806s0010.1
Marpolmon MpmSA2 0283s0150.1
Marpolpol MppBR5 0599s0110.1
Marpolrud Mp8g18450.1
Cycpan CYCAS 020742
Cycpan CYCAS 020743
Medtru mRNA MtrunA17Chr8g0385081
Orysat LOC Os02g34580.1.MSUv7.0
Orysat LOC Os01g65000.1.MSUv7.0
Ambtri evm 27.model.AmTr scaffold00017.61
Orysat LOC Os03g62200.1.MSUv7.0
Cycpan CYCAS 028968
Selmoe rna XM 024683398.1
Selmoe rna XM 024675075.1
Orysat LOC Os11g01410.1.MSUv7.0
Orysat LOC Os12g01420.1.MSUv7.0
Ambtri evm 27.model.AmTr scaffold00007.138
Medtru mRNA MtrunA17Chr4g0007621
Medtru mRNA MtrunA17Chr7g0243071
Medtru mRNA MtrunA17Chr4g0007611
Aratha AT2G38290.1.TAIR10
Orysat LOC Os05g39240.1.MSUv7.0
Orysat LOC Os01g42740.1.MSUv7.0
Orysat LOC Os01g61550.1.MSUv7.0
Orysat LOC Os01g61510.1.MSUv7.0
Marpal utg000013g0032291
Luncru64630 Lc composite 23199.mRNA1
Luncru64630 Lc composite 00045.mRNA1
Cycpan CYCAS 007529
Ambtri evm 27.model.AmTr scaffold00001.151
Medtru mRNA MtrunA17Chr8g0371171
Medtru mRNA MtrunA17Chr8g0371161
Ambtri evm 27.model.AmTr scaffold00001.150
Orysat LOC Os03g53780.1.MSUv7.0
Medtru mRNA MtrunA17Chr1g0163571
Ambtri evm 27.model.AmTr scaffold00002.555
Medtru mRNA MtrunA17Chr7g0274111
Juninf g13309.t1
Calfis g15968.t1
Calfis g36226.t2
Juninf g13306.t1
Calfis g36228.t1
Calfis g15970.t1
Juninf g13308.t1
Calfis g15969.t1
Calfis g36227.t1
Takep Tle1c06872.1
Takep Tle1c003503.1
Sphfal Sphfalx11G108300.1.v1.1
Sphfal Sphfalx10G107000.1.v1.1
Sphfal Sphfalx15G033500.1.v1.1
Sphfal Sphfalx10G107100.1.v1.1
Sphfal Sphfalx04G133400.1.v1.1
Sphfal Sphfalx11G108200.1.v1.1
Takep Tle1c00119.1
Sphfal Sphfalx05G049500.1.v1.1
Phypat Pp12h420.t1
Cerpur JACMSA CepurR40.4G103600.1
Cerpur JACMSA CepurR40.4G107800.1
Fonant Foant 00712 RA
Hypcur jg22718.t1
Cerpur JACMSA CepurR40.1G269800.1
Fonant Foant 07846 RA
Entsed jg11733.t1
Hypcur jg11636.t1
Cerpur JACMSA CepurR40.11G017200.1
Phypat Pp22h197.t1
Phypat Pp19h076.t1
Phypat Pp25h802.t1
Phypat Pp16h092.t1
Cerpur JACMSA CepurR40.5G072600.1
Fonant Foant 19939 RA
Entsed jg24930.t1
Hypcur jg7733.t1
Sphfal Sphfalx04G075400.1.v1.1
Cerpur JACMSA CepurR40.2G150300.1
Phypat Pp10h415.t1
Phypat Pp14h566.t1
Phypat Pp02h636.t1
Cerpur JACMSA CepurR40.6G135200.1
Hypcur jg2371.t1
Entsed jg6615.t1
Hypcur jg20644.t1
Fonant Foant 14246 RA
Entsed jg2777.t1
Cerpur JACMSA CepurR40.5G061600.1
Phypat Pp05h515.t1
Phypat Pp06h630.t1
Phypat Pp16h771.t1
Cerpur JACMSA CepurR40.5G038400.1
Entsed jg25243.t1
Hypcur jg7520.t1
Fonant Foant 02694 RA
Hypcur jg7526.t1
Entsed jg25247.t1

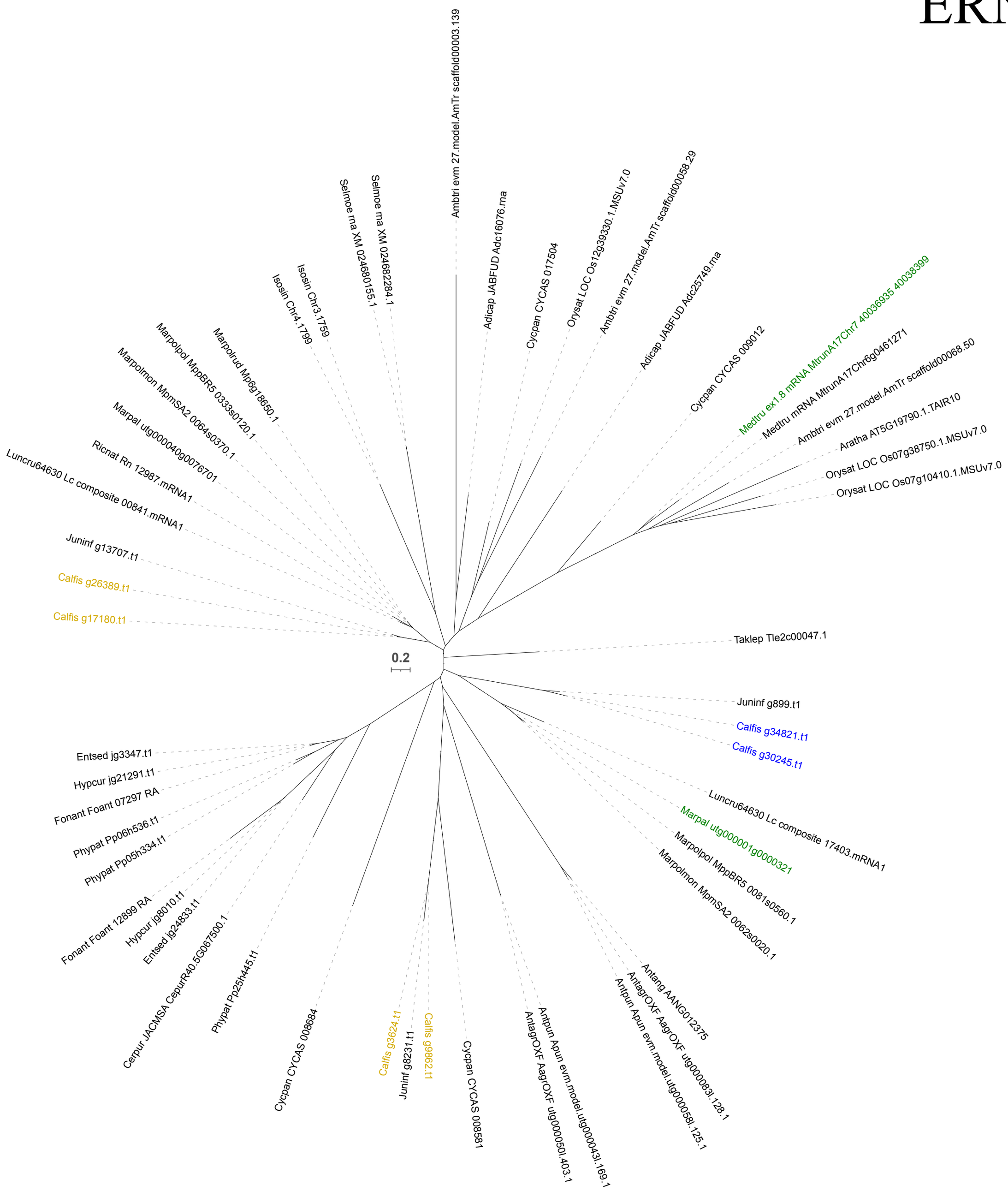
# CCaMK


Onysat LOC Os05g41090.1.MSUv7.0
Medtru mRNA MtrunA17Chr8g0355131
Ambtri evm 27.model.AmTr scaffold00101.38
Taklep Tle1c02117.1
Cerpur JACMSA CepurR40.11G098700.1
Fonant Foant 07025 RA
Hypcur jg11428.t1
Entsed jg11828.t1
Phypat Pp19h733.t1
Phypat Pp21h602.t1
Cerpur JACMSA CepurR40.8G172400.1
Fonant Foant 12058 RA
Hypcur jg3578.t1
Entsed jg20601.t1
Adicap JABFUD Adc08400.rna
0.2
Marpolpol MppBR5 0011s0140.1
Marpolmon MpmSA2 0225s0140.1
Luncru64630 Lc composite 03534.mRNA1
Marpal utg000137g0173321
Calfis g4443.t1
Calfis g10556.t1
Juninf g4536.t1
Antang AANG013471
AntagrOXF AagrOXF utg000006l.689.1
Antpun Apun evm.model.utg0001885.249.1
Selmoe rna XM 024669949.1
Selmoe rna XM 024682220.1

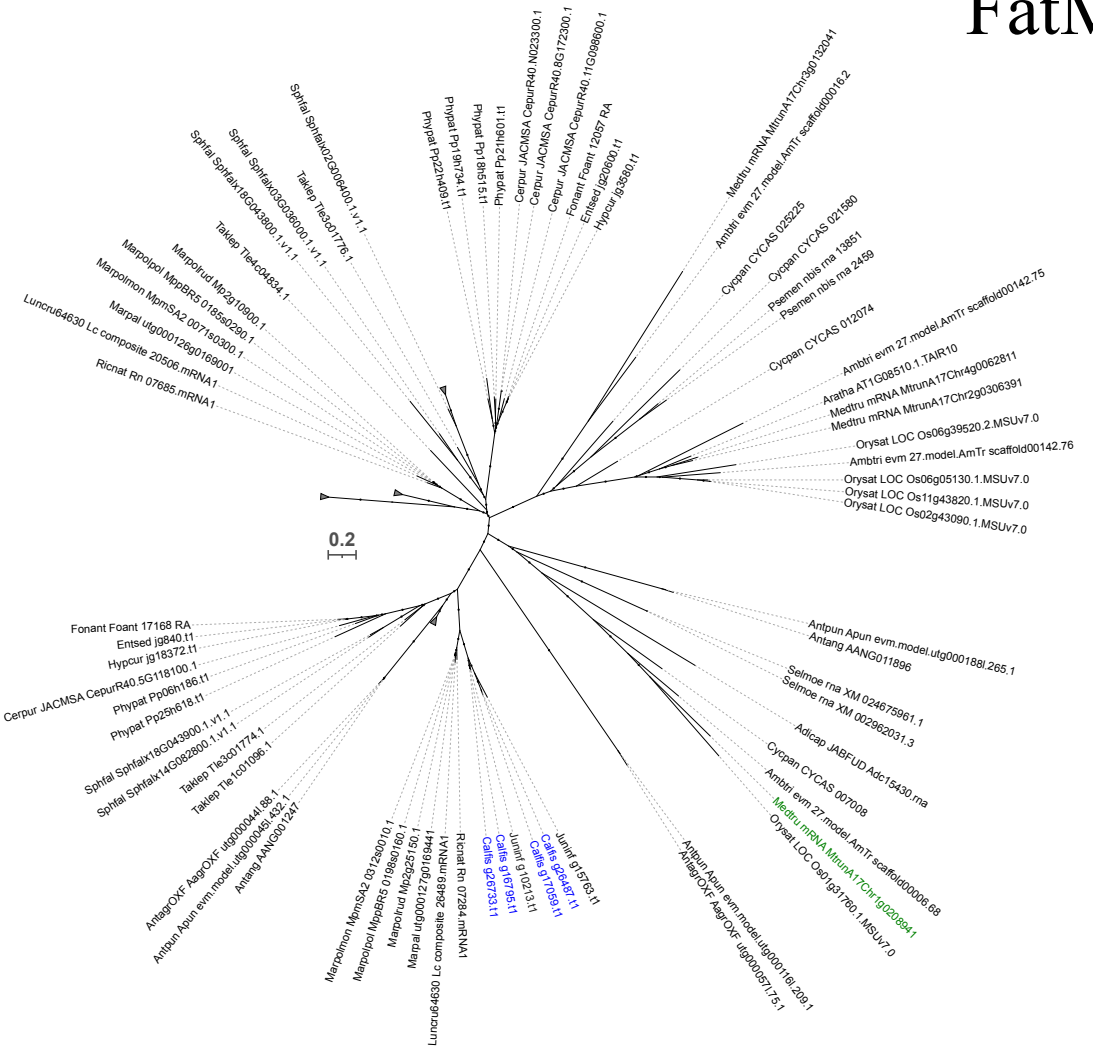
# CYCLOPS


Selmoe rna XM 024672589.1
Selmoe rna XM 024669168.1
Marpal utg000051g0091871
Juninf g2446.t1
Calfis g35878.t2
Calfis g2310.t1
Adicap JABFUD Adc07570.rna
Cycpan CYCAS 014702
Antang AANG007182
Antpun Apun evm.model.utg000038l.656.1
AntagrOXF AagrOXF utg000009l.69.1
0.2
Ambtri evm 27.model.AmTr scaffold00048.167
Orysat LOC Os06g02520.1.MSUv7.0
Medtru mRNA MtrunA17Chr5g0409341
Medtru mRNA MtrunA17Chr8g0369511
Taklep Tle1c06210.1
Hypcur jg17513.t1
Fonant Foant 14918 RA
Entsed jg286.t1
Cerpur JACMSA CepurR40.12G189200.1
Phypat Pp23h831.t3

ERN

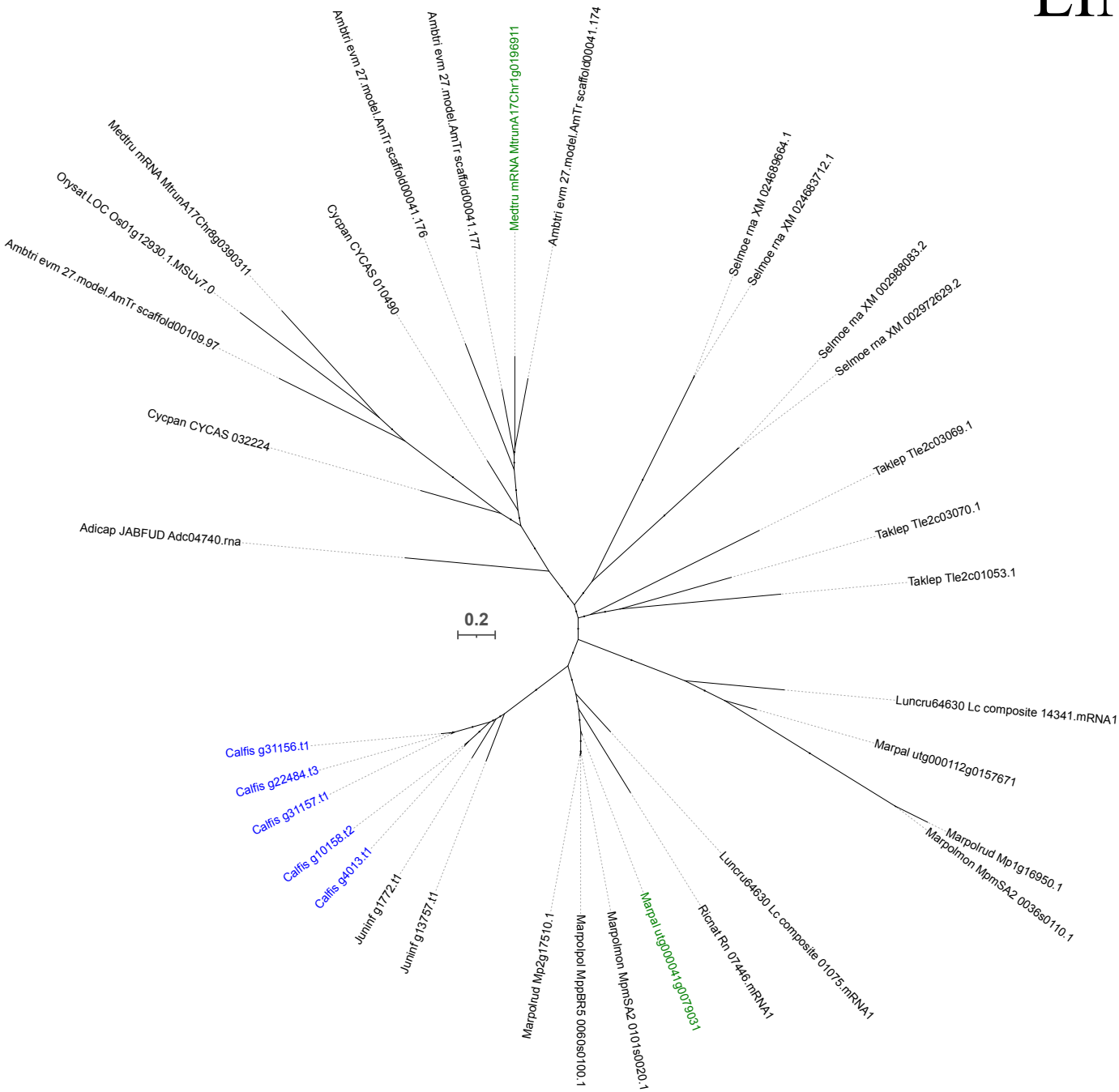

Ambtri evm 27.model.AmTr scaffold00003.139
Adicap JABFUD Adc16076.rna
Cycpan CYCAS 017504
Orysat LOC Os12g39330.1.MSUv7.0
Ambtri evm 27.model.AmTr scaffold00058.29
Adicap JABFUD Adc25749.rna
Cycpan CYCAS 009012
Medtru ex1.8 mRNA MtrunA17Chr7 40036935 40038399
Medtru mRNA MtrunA17Chr6g0461271
Ambtri evm 27.model.AmTr scaffold00068.50
Aratha AT5G19790.1.TAIR10
Orysat LOC Os07g38750.1.MSUv7.0
Orysat LOC Os07g10410.1.MSUv7.0
Taklep Tle2c00047.1
Juninf g899.t1
Calfis g34821.t1
Calfis g30245.t1
Luncru64630 Lc composite 17403.mRNA1
Marpal utg000001g0000321
Marpolpol MppBR5 0081s0560.1
Marpolmon MpmSA2 0062s0020.1
Antang AANG012375
AntagrOXF AagrOXF utg000083l.128.1
Antpun Apun evm.model.utg000058l.125.1
Antpun Apun evm.model.utg000043l.169.1
AntagrOXF AagrOXF utg000050l.403.1
Cycpan CYCAS 008581
Calfis g9862.t1
Juninf g8231.t1
Calfis g3624.t1
Cycpan CYCAS 008684
Phypat Pp25h445.t1
Cerpur JACMSA CepurR40.5G067500.1
Entsed jg24833.t1
Hypcur jg8010.t1
Fonant Foant 12899 RA
Phypat Pp05h334.t1
Phypat Pp06h536.t1
Fonant Foant 07297 RA
Hypcur jg21291.t1
Entsed jg3347.t1
0.2
Calfis g17180.t1
Calfis g26389.t1
Juninf g13707.t1
Luncru64630 Lc composite 00841.mRNA1
Ricnat Rn 12987.mRNA1
Marpal utg000040g0076701
Marpolmon MpmSA2 0064s0370.1
Marpolpol MppBR5 0333s0120.1
Marpolrud Mp6g18650.1
Isosin Cht4.1799
Isosin Chr3.1759
Selmoe rna XM 024680155.1
Selmoe rna XM 024682284.1

# FatM

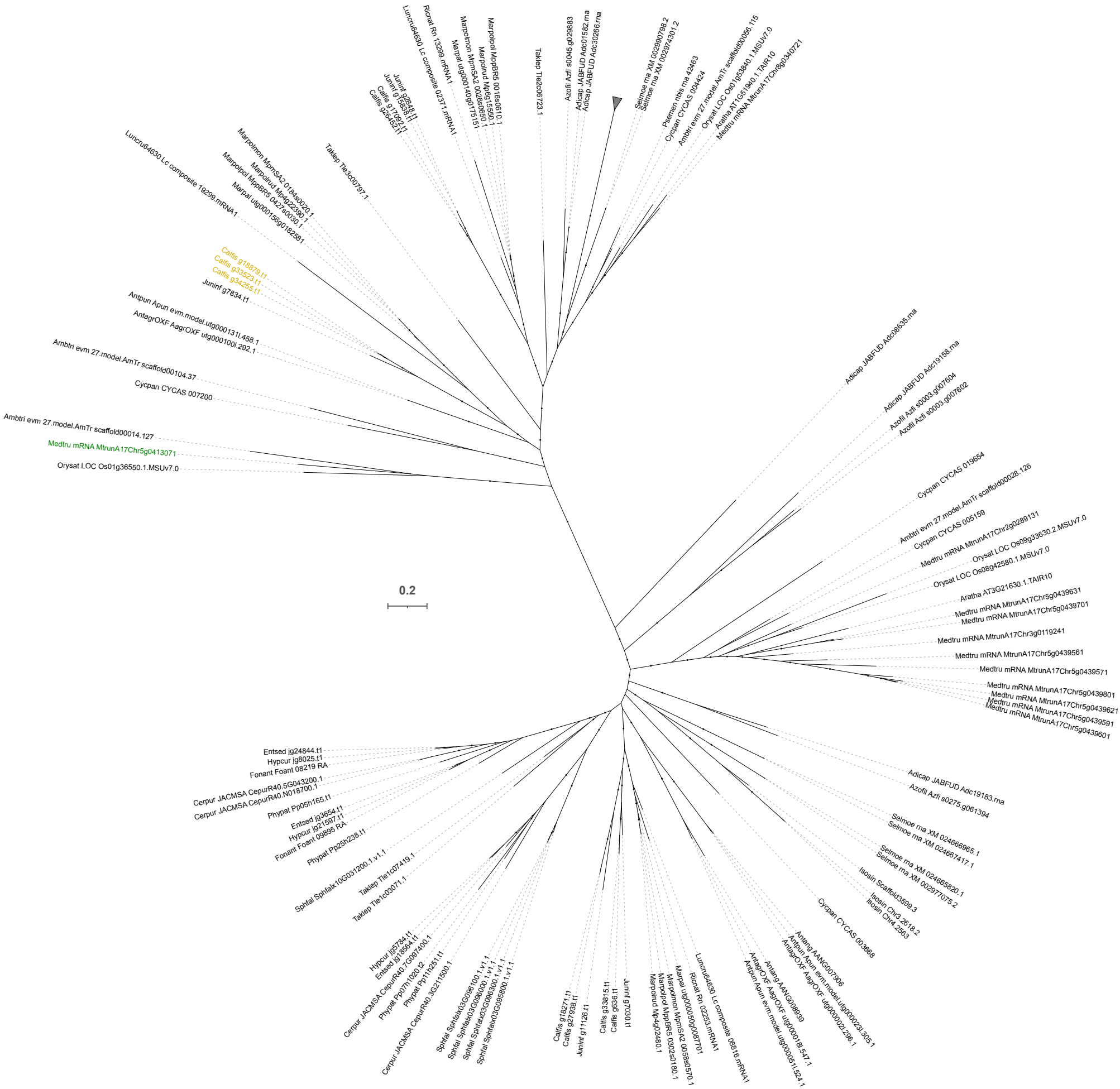


Medtru mRNA MtrunA17Chr3g0132041
Ambtri evm 27.model.AmTr scaffold00016.2
Cycpan CYCAS 025225
Cycpan CYCAS 021580
Psemen nbis rna 13851
Psemen nbis rna 2459
Cycpan CYCAS 012074
Ambtri evm 27.model.AmTr scaffold00142.75
Aratha AT1G08510.1.TAIR10
Medtru mRNA MtrunA17Chr4g0062811
Medtru mRNA MtrunA17Chr2g0306391
Orysat LOC Os06g39520.2.MSUv7.0
Ambtri evm 27.model.AmTr scaffold00142.76
Orysat LOC Os06g05130.1.MSUv7.0
Orysat LOC Os11g43820.1.MSUv7.0
Orysat LOC Os02g43090.1.MSUv7.0
Antpun Apun evm.model.utg000188l.265.1
Antang AANG011896
Selmoe rna XM 024675961.1
Selmoe rna XM 002962031.3
Adicap JABFUD Adc15430.rna
Cycpan CYCAS 007008
Ambtri evm 27.model.AmTr scaffold00006.68
Medtru mRNA MtrunA17Chr1g0208941
Orysat LOC Os01g31760.1.MSUv7.0
Antpun Apun evm.model.utg000116l.209.1
AntagrOXF AagrOXF utg000057l.75.1
Junint g15763.t1
Calfis g26487.t1
Calfis g17059.t1
Junint g10213.t1
Calfis g16795.t1
Calfis g26733.t1
Ricnat Rn 07284.mRNA1
Luncru64630 Lc composite 26489.mRNA1
Marpal utg000127g0169441
Marpolrud Mp2g25150.1
Marpolpol MppBR5 0198s0160.1
Marpolmon MpmSA2 0312s0010.1
Antpun Apun evm.model.utg000045l.432.1
AntagrOXF AagrOXF utg000044l.88.1
Antang AANG001247
Taklep Tle1c01096.1
Taklep Tle3c01774.1
Sphfal Sphfalx14G082800.1.v1.1
Sphfal Sphfalx18G043900.1.v1.1
Phypat Pp25h618.t1
Phypat Pp06h186.t1
Cerpur JACMSA CepurR40.5G118100.1
Hypcur jg18372.t1
Entsed jg840.t1
Fonant Foant 17168 RA
Phypat Pp22h409.t1
Phypat Pp19h734.t1
Phypat Pp18h515.t1
Phypat Pp21h601.t1
Cerpur JACMSA CepurR40.N023300.1
Cerpur JACMSA CepurR40.8G172300.1
Cerpur JACMSA CepurR40.11G098600.1
Fonant Foant 12057 RA
Entsed jg20600.t1
Hypcur jg3580.t1
Sphfal Sphfalx02G006400.1.v1.1
Taklep Tle3c01776.1
Sphfal Sphfalx03G036000.1.v1.1
Sphfal Sphfalx18G043800.1.v1.1
Taklep Tle4c04834.1
Marpolrud Mp2g10900.1
Marpolpol MppBR5 0185s0290.1
Marpolmon MpmSA2 0071s0300.1
Marpal utg000126g0169001
Luncru64630 Lc composite 20506.mRNA1
Ricnat Rn 07685.mRNA1
0.2

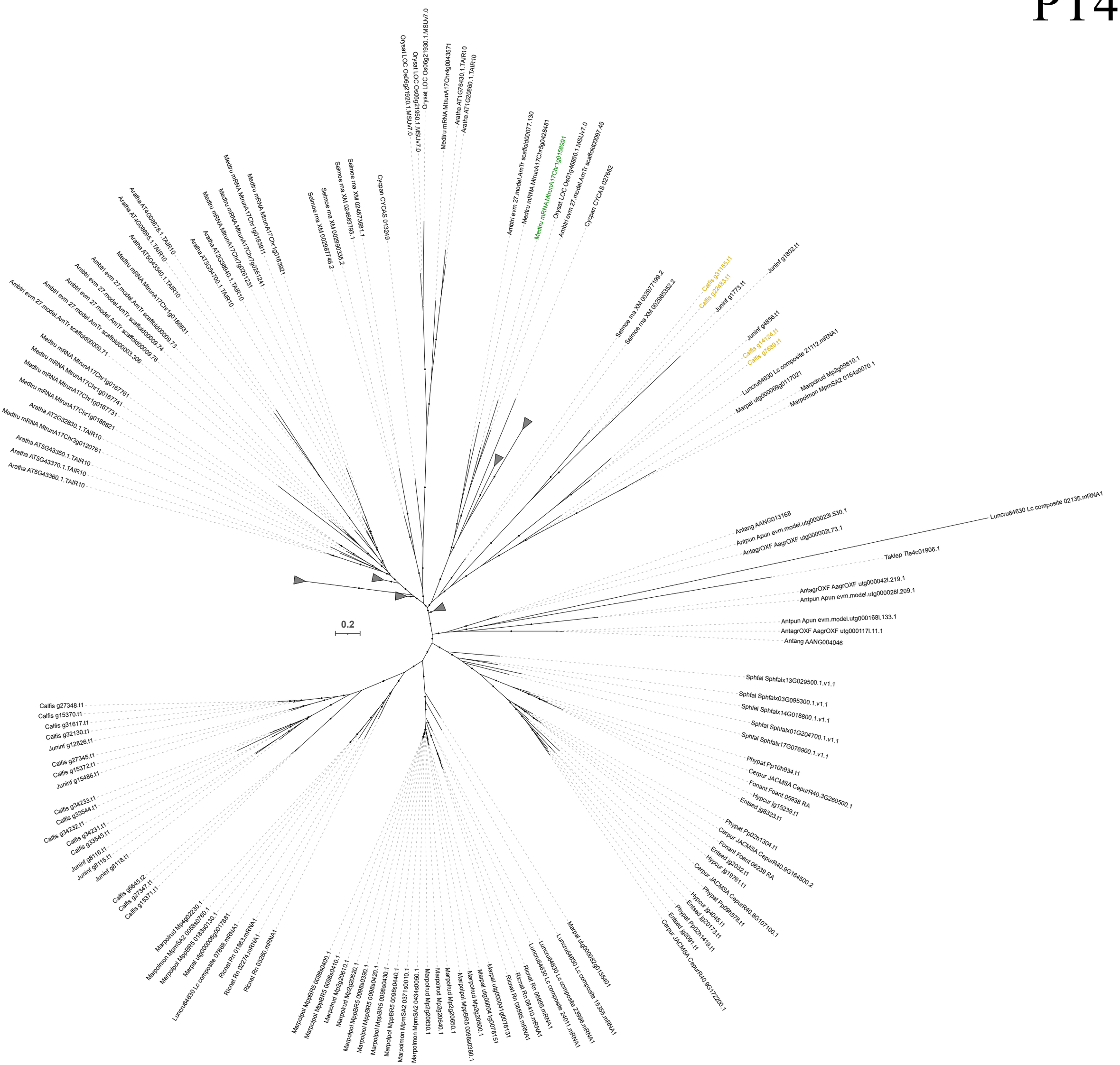
LIN


Ambtri evm 27.model.AmTr scaffold00041.176
Ambtri evm 27.model.AmTr scaffold00041.177
Medtru mRNA MtrunA17Chr1g0196911
Ambtri evm 27.model.AmTr scaffold00041.174
Medtru mRNA MtrunA17Chr8g0390311
Orysat LOC Os01g12930.1 MSUv7.0
Cycpan CYCAS 010490
Ambtri evm 27.model.AmTr scaffold00109.97
Cycpan CYCAS 032224
Adicap JABFUD Adc04740.rna
Selmoe rna XM 024689664.1
Selmoe rna XM 024683712.1
Selmoe rna XM 002988083.2
Selmoe rna XM 002972629.2
Taklep Tle2c03069.1
Taklep Tle2c03070.1
Taklep Tle2c01053.1
0.2
Luncru64630 Lc composite 14341.mRNA1
Marpal utg000112g0157671
Marpolrud Mp1g16950.1
Marpolmon MpmSA2 0036s0110.1
Luncru64630 Lc composite 01075.mRNA1
Richat Rn 07446.mRNA1
Marpal utg000041g0079031
Marpolmon MpmSA2 0101s0020.1
Marpolpol MppBR5 0060s0100.1
Marpolrud Mp2g17510.1
Juninf g13757.t1
Juninf g1772.t1
Calfis g4013.t1
Calfis g10158.t2
Calfis g31157.t1
Calfis g22484.t3
Calfis g31156.t1

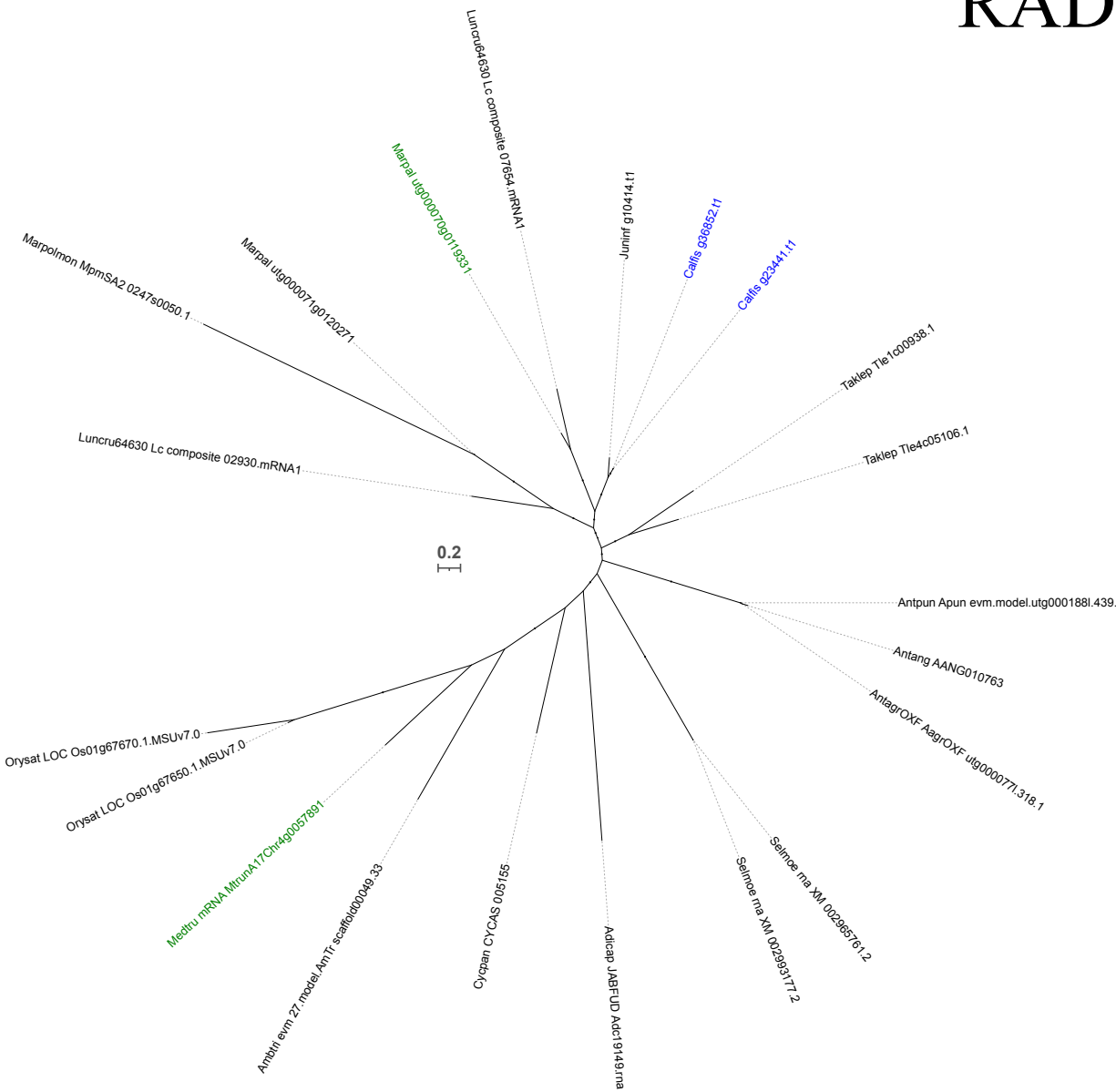
# LYK10/EPR3


0.2
Taklep Tle2c06723.1
Azofil Azfi s0045.g029883
Adicap JABFUD Adc01582.rna
Adicap JABFUD Adc30266.rna
Selmoe rna XM 002990798.2
Selmoe rna XM 002974301.2
Psemen nbis rna 42463
Cycpan CYCAS 004424
Ambtri evm 27.model.AmTr scaffold00056.115
Orysat LOC Os01g53840.1.MSUv7.0
Aratha AT1G51940.1.TAIR10
Medtru mRNA MtrunA17Chr8g0340721
Ricnat Rn 13299.mRNA1
Luncru64630 Lc composite 02371.mRNA1
Marpol utg000140g0175151
Marpolmon MpmSA2 0026s0650.1
Marpolrud Mp4g15550.1
Marpolpol MppBR5 0016s0610.1
Juninf g28481.t1
Juninf g15638.t1
Calfis g17092.t1
Calfis g26452.t1
Taklep Tle3c00797.1
Marpolmon MpmSA2 0184s0020.1
Marpolrud Mp4g22390.1
Marpolpol MppBR5 0427s0030.1
Marpal utg000156g0182581
Luncru64630 Lc composite 19299.mRNA1
Calfis g18879.t1
Calfis g33523.t1
Calfis g34255.t1
Juninf g7834.t1
Antpun Apun evm.model.utg000131l.458.1
AntagrOXF AagrOXF utg000100l.292.1
Ambtri evm 27.model.AmTr scaffold00104.37
Cycpan CYCAS 007200
Ambtri evm 27.model.AmTr scaffold00014.127
Medtru mRNA MtrunA17Chr5g0413071
Orysat LOC Os01g36550.1.MSUv7.0
Adicap JABFUD Adc08635.rna
Adicap JABFUD Adc19158.rna
Azofil Azfi s0003.g007604
Azofil Azfi s0003.g007602
Cycpan CYCAS 019654
Ambtri evm 27.model.AmTr scaffold00028.126
Cycpan CYCAS 005159
Medtru mRNA MtrunA17Chr2g0289131
Orysat LOC Os09g33630.2.MSUv7.0
Orysat LOC Os08g42580.1.MSUv7.0
Aratha AT3G21630.1.TAIR10
Medtru mRNA MtrunA17Chr5g0439631
Medtru mRNA MtrunA17Chr5g0439701
Medtru mRNA MtrunA17Chr3g0119241
Medtru mRNA MtrunA17Chr5g0439561
Medtru mRNA MtrunA17Chr5g0439571
Medtru mRNA MtrunA17Chr5g0439801
Medtru mRNA MtrunA17Chr5g0439621
Medtru mRNA MtrunA17Chr5g0439591
Medtru mRNA MtrunA17Chr5g0439601
Entsed jg24844.t1
Hypcur jg8025.t1
Fonant Foant 08219 RA
Cerpur JACMSA CepurR40.5G043200.1
Cerpur JACMSA CepurR40.N018700.1
Phypat Pp05h165.t1
Entsed jg3654.t1
Hypcur jg21597.t1
Fonant Foant 09895 RA
Phypat Pp25h238.t1
Sphfal Sphfalx10G031200.1.v1.1
Taklep Tle1c07419.1
Taklep Tle1c03071.1
Hypcur jg5784.t1
Entsed jg18564.t1
Cerpur JACMSA CepurR40.7G097400.1
Phypat Pp07h1020.t2
Phypat Pp11h251.t1
Cerpur JACMSA CepurR40.3G211500.1
Sphfal Sphfalx03G096100.1.v1.1
Sphfal Sphfalx03G096000.1.v1.1
Sphfal Sphfalx03G096300.1.v1.1
Sphfal Sphfalx03G095800.1.v1.1
Calfis g18271.t1
Calfis g27938.t1
Juninf g11126.t1
Calfis g33815.t1
Calfis g636.t1
Juninf g7030.t1
Marpolpol MppBR5 0302s0180.1
Marpolmon MpmSA2 0058s0570.1
Marpolrud Mp4g02480.1
Marpal utg000059g0087701
Ricnat Rn 02253.mRNA1
Luncru64630 Lc composite 08816.mRNA1
AntagrOXF AagrOXF utg000018l.547.1
Antpun Apun evm.model.utg000051l.524.1
Antang AANG008939
AntagrOXF AagrOXF utg000021.296.1
Antang AANG007906
Antpun Apun evm.model.utg000023l.305.1
Cycpan CYCAS 003868
Isosin Chr3.2618.2
Isosin Chr4.2563
Isosin Scaffold3599.3
Selmoe rna XM 024665820.1
Selmoe rna XM 002977075.2
Selmoe rna XM 024667417.1
Selmoe rna XM 024666965.1
Adicap JABFUD Adc19183.rna
Azofil Azfi s0275.g061394

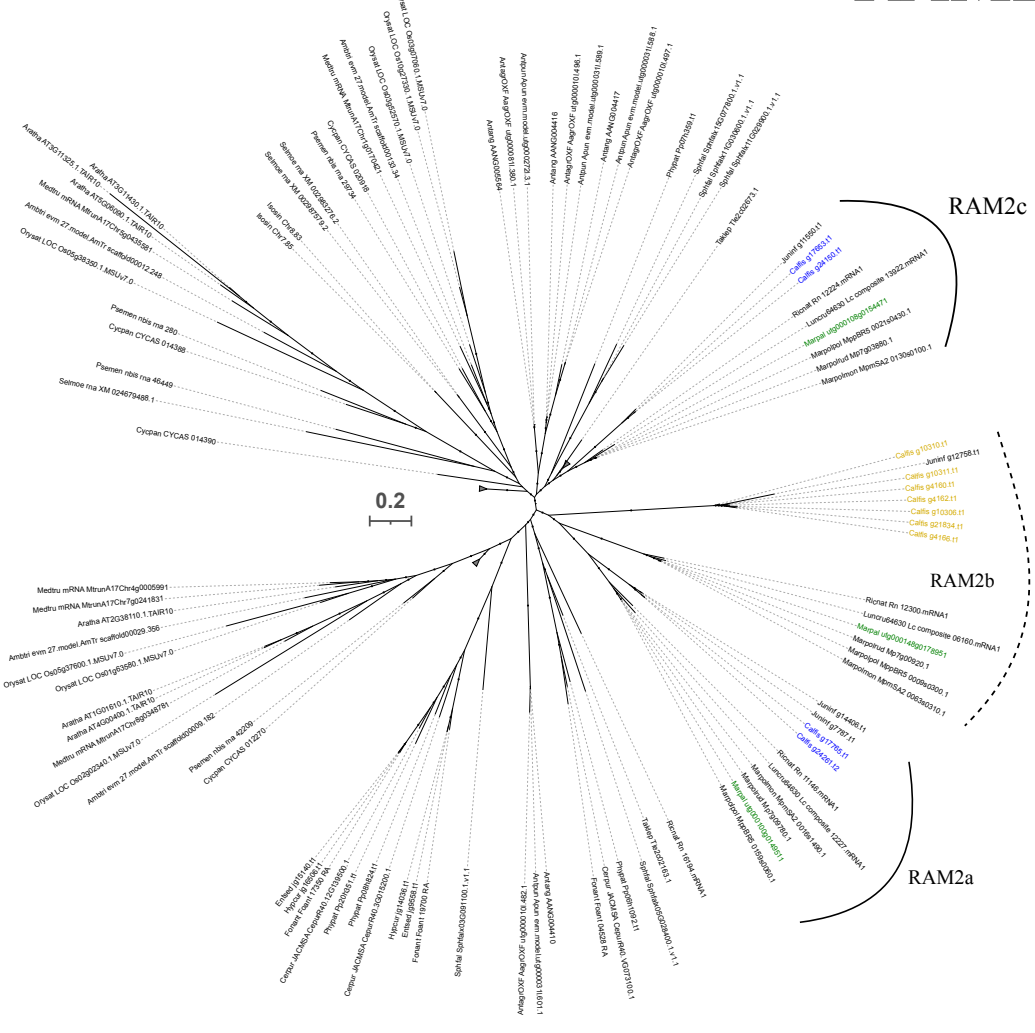
# PT4

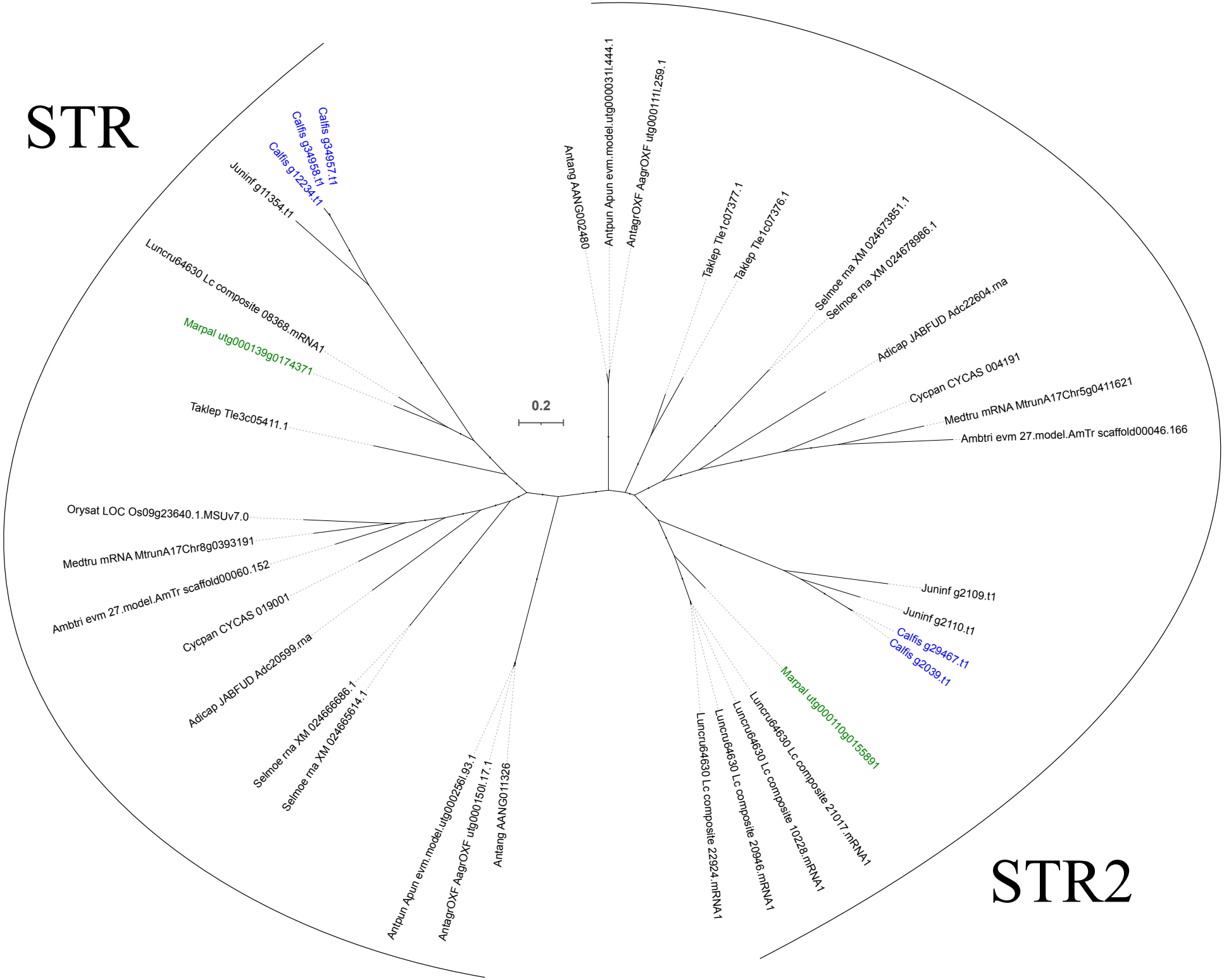

Orysat LOC Os06g21930.1.MSUv7.0
Medtru mRNA MtrunA17Chr4g0043571
Aratha AT1G76430.1.TAIR10
Aratha AT1G20860.1.TAIR10
Ambtri evm 27.model.AmTr.scaffold00077.130
Medtru mRNA MtrunA17Chr5g0428481
Medtru mRNA MtrunA17Chr1g0158991
Orysat LOC Os01g46860.1.MSUv7.0
Ambtri evm 27.model.AmTr.scaffold00097.45
Cycpan CYCAS 027682
Selmoe rna XM 024673681.1
Selmoe rna XM 002987746.2
Selmoe rna XM 002990335.2
Selmoe rna XM 024663793.1
Cycpan CYCAS 013249
Medtru mRNA MtrunA17Chr1g0183921
Medtru mRNA MtrunA17Chr1g0183911
Medtru mRNA MtrunA17Chr7g0261241
Medtru mRNA MtrunA17Chr7g0261231
Aratha AT2G39840.1.TAIR10
Aratha AT3G54700.1.TAIR10
Aratha AT4G08878.1.TAIR10
Aratha AT4G08895.1.TAIR10
Aratha AT5G43340.1.TAIR10
Medtru mRNA MtrunA17Chr1g0186831
Ambtri evm 27.model.AmTr.scaffold00009.73
Ambtri evm 27.model.AmTr.scaffold00009.74
Ambtri evm 27.model.AmTr.scaffold00009.76
Ambtri evm 27.model.AmTr.scaffold00003.306
Ambtri evm 27.model.AmTr.scaffold00009.71
Medtru mRNA MtrunA17Chr1g0167761
Medtru mRNA MtrunA17Chr1g0167741
Medtru mRNA MtrunA17Chr1g0167731
Medtru mRNA MtrunA17Chr1g0186821
Aratha AT2G32830.1.TAIR10
Medtru mRNA MtrunA17Chr3g0120761
Aratha AT5G43350.1.TAIR10
Aratha AT5G43370.1.TAIR10
Aratha AT5G43360.1.TAIR10
Selmoe rna XM 002977199.2
Selmoe rna XM 002965352.2
Calfis g31155.t1
Calfis g22483.t1
Juninf g1773.t1
Juninf g1802.t1
Juninf g4856.t1
Calfis g14124.t1
Calfis g7689.t1
Luncru64630 Lc composite 21112.mRNA1
Marpolrud Mp2g09810.1
Marpai utg000069g0117021
Marpolmon MpmSA2 0164s0070.1
Antang AANG013168
Antpun Apun evm.model.utg000023l.530.1
AntagrOXF AagrOXF utg000002l.73.1
Luncru64630 Lc composite 02135.mRNA1
Taklep Tle4c01906.1
AntagrOXF AagrOXF utg000042l.219.1
Antpun Apun evm.model.utg000028l.209.1
Antpun Apun evm.model.utg000168l.133.1
AntagrOXF AagrOXF utg000117l.11.1
Antang AANG004046
0.2
Sphfal Sphfalx13G029500.1.v1.1
Sphfal Sphfalx03G095300.1.v1.1
Sphfal Sphfalx14G018800.1.v1.1
Sphfal Sphfalx01G204700.1.v1.1
Sphfal Sphfalx17G076900.1.v1.1
Phypat Pp10h934.t1
Cerpur JACMSA CepurR40.3G260500.1
Fonant Foant 05938 RA
Hypcur jg15239.t1
Entsed jg8323.t1
Phypat Pp02h1304.t1
Cerpur JACMSA CepurR40.9G164500.2
Fonant Foant 06239 RA
Entsed jg2032.t1
Hypcur jg19761.t1
Cerpur JACMSA CepurR40.8G107100.1
Phypat Pp09h578.t1
Hypcur jg4045.t1
Entsed jg2173.t1
Phypat Pp02h1419.t1
Entsed jg2091.t1
Cerpur JACMSA CepurR40.9G172200.1
Marpai utg000092g0135401
Luncru64630 Lc composite 15355.mRNA1
Luncru64630 Lc composite 23996.mRNA1
Luncru64630 Lc composite 24011.mRNA1
Ricnat Rn 06995.mRNA1
Ricnat Rn 08410.mRNA1
Ricnat Rn 08595.mRNA1
Marpai utg000041g0078131
Marpolrud Mp2g20600.1
Marpolpol MppBR5 0098s0380.1
Marpolrud Mp2g20650.1
Marpolrud Mp2g20640.1
Marpolrud Mp2g20630.1
Marpolmon MpmSA2 0434s0050.1
Marpolmon MpmSA2 0371s0010.1
Marpolpol MppBR5 0098s0440.1
Marpolpol MppBR5 0098s0430.1
Marpolpol MppBR5 0098s0420.1
Marpolpol MppBR5 0098s0390.1
Marpolrud Mp2g20620.1
Marpolrud Mp2g20610.1
Marpolpol MppBR5 0098s0410.1
Marpolpol MppBR5 0098s0400.1
Calfis g27348.t1
Calfis g15370.t1
Calfis g31617.t1
Calfis g32130.t1
Juninf g12826.t1
Calfis g27345.t1
Calfis g15372.t1
Juninf g15486.t1
Calfis g34233.t1
Calfis g33544.t1
Calfis g34232.t1
Calfis g34231.t1
Calfis g33545.t1
Juninf g8116.t1
Juninf g8115.t1
Juninf g8118.t1
Calfis g6645.t2
Calfis g27347.t1
Calfis g15371.t1
Marpolrud Mp4g02230.1
Marpolmon MpmSA2 0058s0760.1
Marpolpol MppBR5 0183s0130.1
Marpai utg000008g0017881
Luncru64630 Lc composite 07868.mRNA1
Ricnat Rn 01863.mRNA1
Ricnat Rn 02274.mRNA1
Ricnat Rn 03280.mRNA1

# RAD1

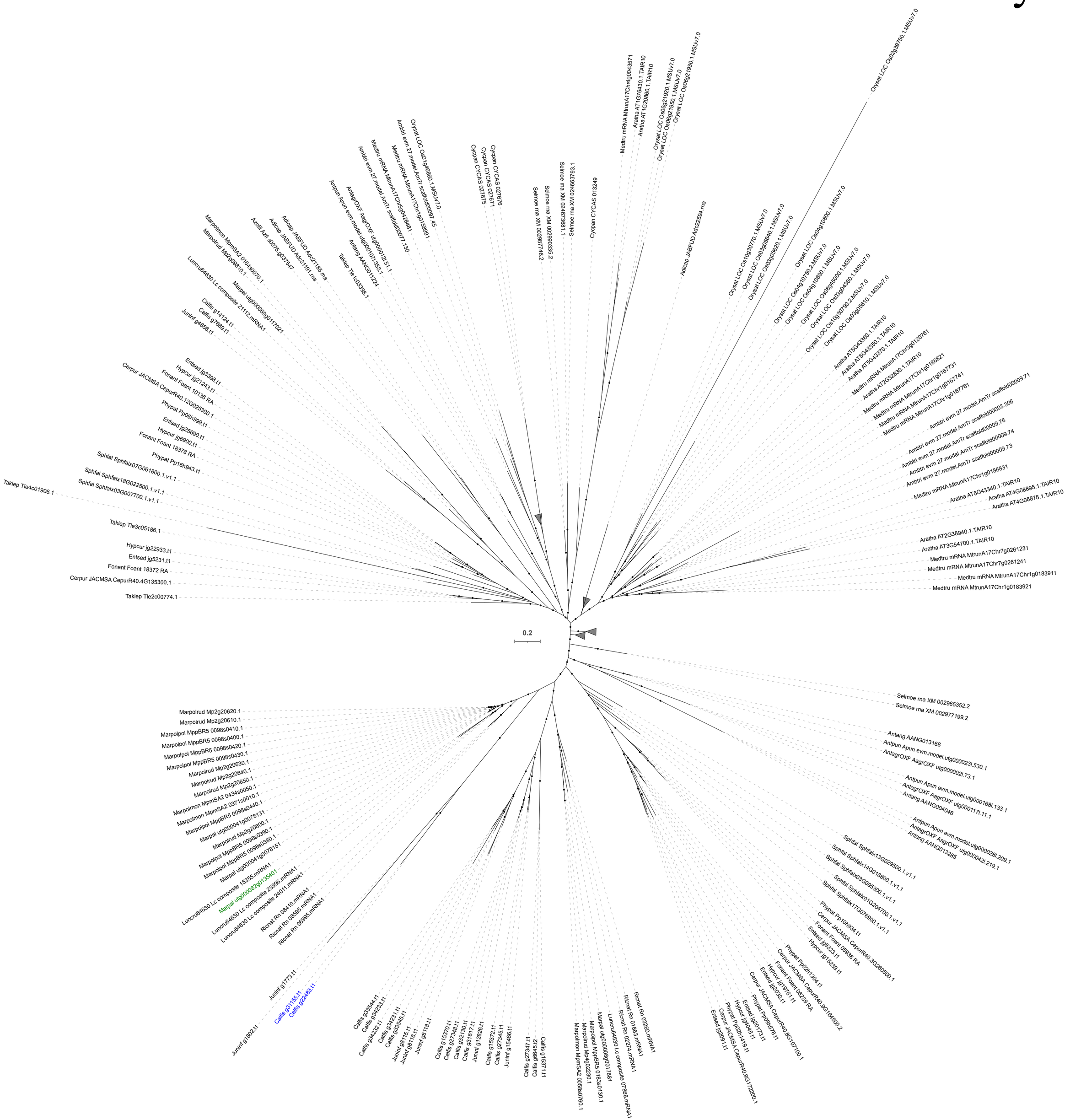

Luncru64630 Lc composite 07654.mRNA1
Marpal utg000079g0119331
Juninf g10414.t1
Callis g36852.t1
Callis g23441.t1
Marpolmon MpmSA2 0247s0050.1
Marpal utg000071g0120271
Taklep Tle1c00938.1
Luncru64630 Lc composite 02930.mRNA1
Taklep Tle4c05106.1
0.2
Antpun Apun evm.model.utg000188l.439.1
Antang AANG010763
AntagrOXF AagrOXF utg000077l.318.1
Orysat LOC Os01g67670.1.MSUv7.0
Orysat LOC Os01g67650.1.MSUv7.0
Medtru mRNA MtrunA17Chr4g0057891
Ambtri evm 27.model.AmTr scaffold00049.33
Cycpan CYCAS 005155
Adicap JABFUD Adc19149.rna
Selmoe ma XM 002993777.2
Selmoe ma XM 002965761.2

# RAM2

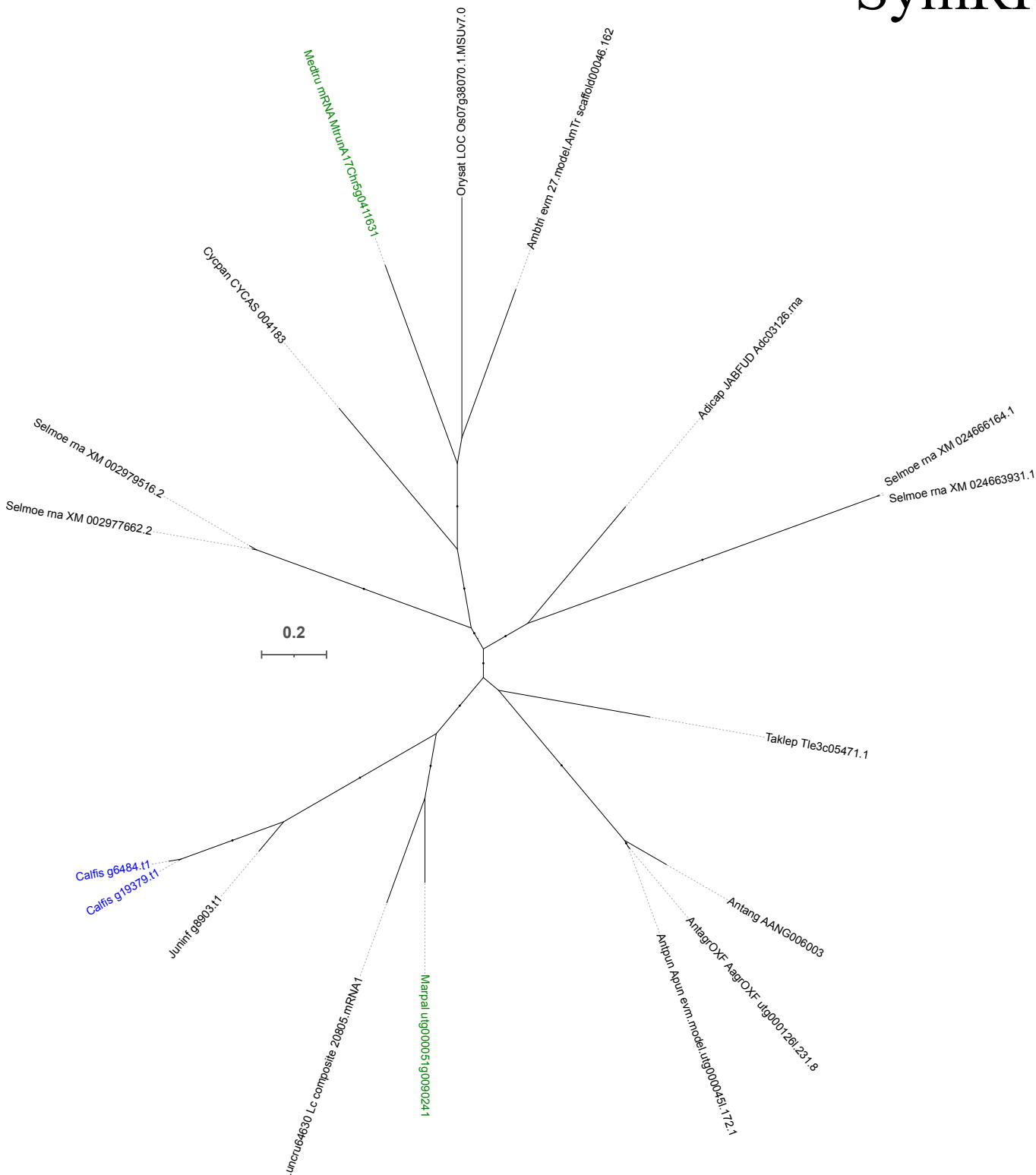



STR
STR2
0.2
Antang AANG002480
Antpun Apun evm.model.utg000031l.444.1
AntagrOXF AagrOXF utg000111l.259.1
Taklep Tle1c07377.1
Taklep Tle1c07376.1
Selmoe rna XM 024673851.1
Selmoe rna XM 024678986.1
Adicap JABFUD Adc22604.rna
Cycpan CYCAS 004191
Medtru mRNA MtrunA17Chr5g0411621
Ambtri evm 27.model.AmTr scaffold00046.166
Juninf g2109.t1
Juninf g2110.t1
Calfis g29467.t1
Calfis g2039.t1
Marpal utg000110g0155891
Luncru64630 Lc composite 21017.mRNA1
Luncru64630 Lc composite 10228.mRNA1
Luncru64630 Lc composite 20946.mRNA1
Luncru64630 Lc composite 22924.mRNA1
Antang AANG011326
AntagrOXF AagrOXF utg000150l.17.1
Antpun Apun evm.model.utg000256l.93.1
Selmoe rna XM 024665614.1
Selmoe rna XM 024666686.1
Adicap JABFUD Adc20599.rna
Cycpan CYCAS 019001
Ambtri evm 27.model.AmTr scaffold00060.152
Medtru mRNA MtrunA17Chr8g0393191
Orysat LOC Os09g23640.1.MSUv7.0
Taklep Tle3c05411.1
Marpal utg000139g0174371
Luncru64630 Lc composite 08368.mRNA1
Juninf g11354.t1
Calfis g12234.t1
Calfis g34958.t1
Calfis g34957.t1

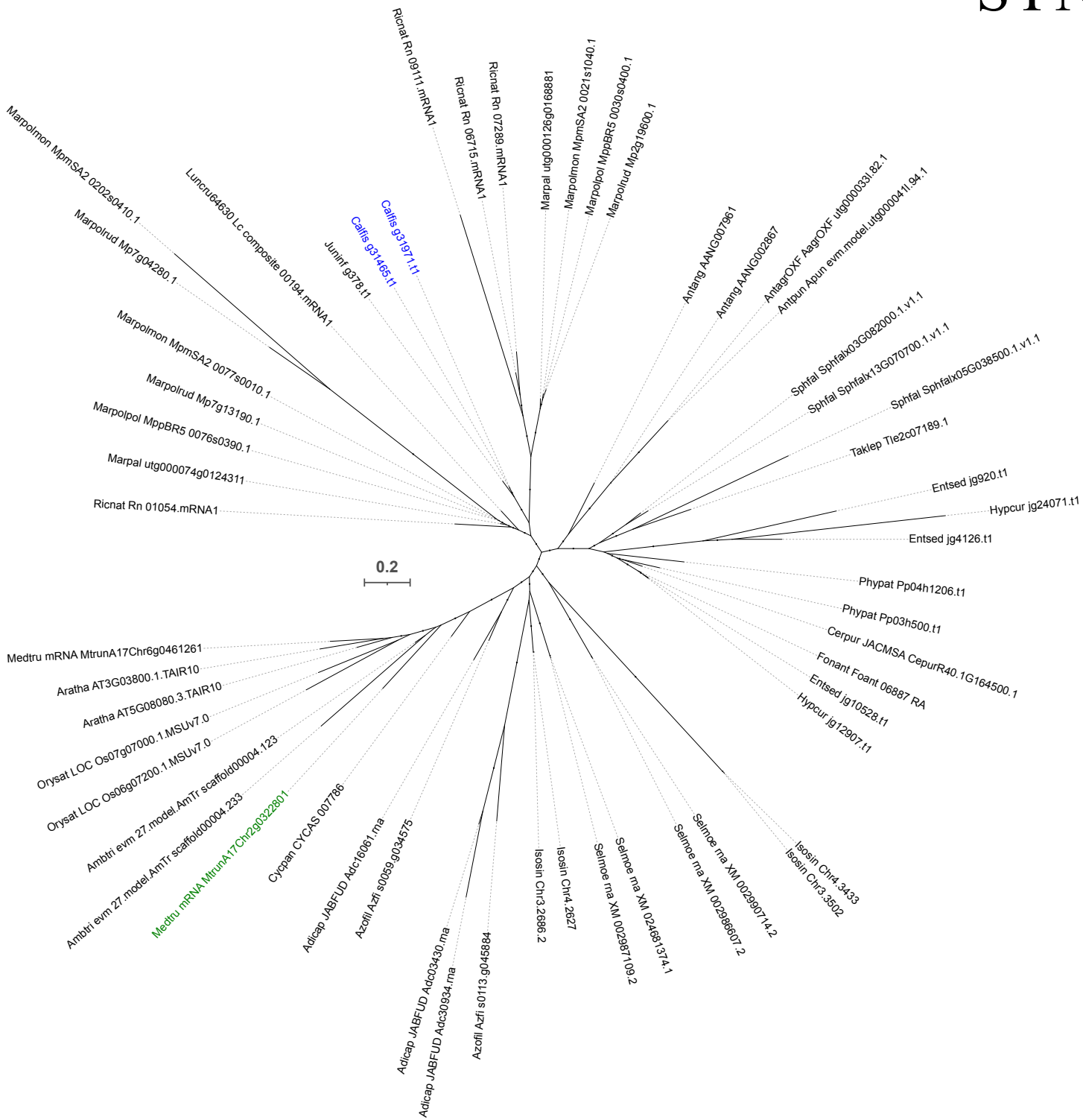
# SymPT


0.2
Orysat LOC Os02g39750.1.MSUv7.0
Medtru mRNA MtrunA17Chr4g0043571
Aratha AT1G76430.1.TAIR10
Aratha AT1G20860.1.TAIR10
Orysat LOC Os06g21920.1.MSUv7.0
Orysat LOC Os06g21950.1.MSUv7.0
Orysat LOC Os06g21930.1.MSUv7.0
Cycpan CYCAS 013249
Adicap JABFUD Adc22594.rna
Selmoe rna XM 024663793.1
Selmoe rna XM 024693681.1
Selmoe rna XM 002990335.2
Selmoe rna XM 002987746.2
Cycpan CYCAS 027676
Cycpan CYCAS 027671
Cycpan CYCAS 027675
Orysat LOC Os01g46660.1.MSUv7.0
Ambtri evm.27.model.AmTr.scaffold00097.45
Medtru mRNA MtrunA17Chr1g0158991
Medtru mRNA MtrunA17Chr5g0426481
Ambtri evm.27.model.AmTr.scaffold00077.190
AntagrOXF AagrOXF utg000001l.51.1
Antpun Apun evm.model.utg000107l.353.1
Antang AANG011224
Taklep Tle1c03398.1
Adicap JABFUD Adc21185.rna
Adicap JABFUD Adc21191.rna
Azofil Azfi s0075.g037547
Marpolmon MpmSA2 0164s0070.1
Marpolrud Mp2g09810.1
Marpal utg000069g0117021
Luncru64630 Lc composite 21112.mRNA1
Callfis g14124.t1
Callfis g7689.t1
Juninf g4856.t1
Entsed jg3398.t1
Hypcur jg21243.t1
Fonant Foant 10136 RA
Cerpur JACMSA CepurR40.12G025300.1
Phypat Pp06h999.t1
Entsed jg25690.t1
Hypcur jg6900.t1
Fonant Foant 18378 RA
Phypat Pp16h943.t1
Sphfal Sphfalx07G061800.1.v1.1
Sphfal Sphfalx18G022500.1.v1.1
Sphfal Sphfalx03G007700.1.v1.1
Taklep Tle4c01906.1
Taklep Tle3c05186.1
Hypcur jg22933.t1
Entsed jg5231.t1
Fonant Foant 18372 RA
Cerpur JACMSA CepurR40.4G135300.1
Taklep Tle2c00774.1
Orysat LOC Os10g30770.1.MSUv7.0
Orysat LOC Os03g05640.1.MSUv7.0
Orysat LOC Os03g05620.1.MSUv7.0
Orysat LOC Os04g10600.1.MSUv7.0
Orysat LOC Os04g10750.2.MSUv7.0
Orysat LOC Os04g10690.1.MSUv7.0
Orysat LOC Os08g45000.1.MSUv7.0
Orysat LOC Os03g04360.1.MSUv7.0
Orysat LOC Os10g30790.2.MSUv7.0
Orysat LOC Os03g05610.1.MSUv7.0
Aratha AT5G43360.1.TAIR10
Aratha AT5G43350.1.TAIR10
Aratha AT5G43370.1.TAIR10
Medtru mRNA MtrunA17Chr3g0120761
Aratha AT2G32830.1.TAIR10
Medtru mRNA MtrunA17Chr1g0186821
Medtru mRNA MtrunA17Chr1g0167731
Medtru mRNA MtrunA17Chr1g0167741
Medtru mRNA MtrunA17Chr1g0167761
Ambtri evm.27.model.AmTr.scaffold00009.71
Ambtri evm.27.model.AmTr.scaffold00003.306
Ambtri evm.27.model.AmTr.scaffold00009.76
Ambtri evm.27.model.AmTr.scaffold00009.74
Ambtri evm.27.model.AmTr.scaffold00009.73
Medtru mRNA MtrunA17Chr1g0186831
Aratha AT5G43340.1.TAIR10
Aratha AT4G08895.1.TAIR10
Aratha AT4G08878.1.TAIR10
Aratha AT2G38940.1.TAIR10
Aratha AT3G54700.1.TAIR10
Medtru mRNA MtrunA17Chr7g0261231
Medtru mRNA MtrunA17Chr7g0261241
Medtru mRNA MtrunA17Chr1g0183911
Medtru mRNA MtrunA17Chr1g0183921
Selmoe rna XM 002965352.2
Selmoe rna XM 002977199.2
Antang AANG013168
Antpun Apun evm.model.utg000023l.530.1
AntagrOXF AagrOXF utg000002l.73.1
Antpun Apun evm.model.utg000168l.133.1
AntagrOXF AagrOXF utg000117l.11.1
Antang AANG004046
Antpun Apun evm.model.utg000028l.209.1
AntagrOXF AagrOXF utg000042l.219.1
Antang AANG013285
Sphfal Sphfalx13G029500.1.v1.1
Sphfal Sphfalx14G018800.1.v1.1
Sphfal Sphfalx03G095300.1.v1.1
Sphfal Sphfalx01G204700.1.v1.1
Sphfal Sphfalx17G076900.1.v1.1
Phypat Pp10h934.t1
Cerpur JACMSA CepurR40.3G260500.1
Fonant Foant 05938 RA
Entsed jg8323.t1
Hypcur jg15239.t1
Phypat Pp02h1304.t1
Cerpur JACMSA CepurR40.9G164500.2
Fonant Foant 06239 RA
Hypcur jg19761.t1
Entsed jg2032.t1
Cerpur JACMSA CepurR40.8G107100.1
Phypat Pp09h578.t1
Entsed jg20173.t1
Hypcur jg9045.t1
Phypat Pp02h1419.t1
Cerpur JACMSA CepurR40.9G172200.1
Entsed jg2091.t1
Marpolrud Mp2g20620.1
Marpolrud Mp2g20610.1
Marpolpol MppBR5 0098s0410.1
Marpolpol MppBR5 0098s0400.1
Marpolpol MppBR5 0098s0420.1
Marpolpol MppBR5 0098s0430.1
Marpolrud Mp2g20630.1
Marpolrud Mp2g20640.1
Marpolrud Mp2g20650.1
Marpolmon MpmSA2 0434s0050.1
Marpolmon MpmSA2 0371s0010.1
Marpolpol MppBR5 0098s0440.1
Marpal utg000041g0078131
Marpolrud Mp2g20600.1
Marpolpol MppBR5 0098s0390.1
Marpolpol MppBR5 0098s0380.1
Marpal utg000041g0078151
Luncru64630 Lc composite 15355.mRNA1
Marpal utg000082g0135401
Luncru64630 Lc composite 23996.mRNA1
Luncru64630 Lc composite 24011.mRNA1
Ricnat Rn 08410.mRNA1
Ricnat Rn 08595.mRNA1
Ricnat Rn 06995.mRNA1
Juninf g1773.t1
Callfis g31155.t1
Callfis g22483.t1
Juninf g1802.t1
Callfis g33544.t1
Callfis g34232.t1
Callfis g34232.t1
Callfis g34231.t1
Callfis g33545.t1
Juninf g8115.t1
Juninf g8116.t1
Juninf g8118.t1
Callfis g15370.t1
Callfis g27348.t1
Callfis g32130.t1
Callfis g31617.t1
Juninf g12826.t1
Callfis g15372.t1
Callfis g27345.t1
Juninf g15486.t1
Callfis g27347.t1
Callfis g6645.t2
Callfis g15371.t1
Marpolmon MpmSA2 0058s0760.1
Marpolrud Mp4g02230.1
Marpolpol MppBR5 0183s0130.1
Marpal utg000008g0017881
Luncru64630 Lc composite 07868.mRNA1
Ricnat Rn 01863.mRNA1
Ricnat Rn 02274.mRNA1
Ricnat Rn 03260.mRNA1

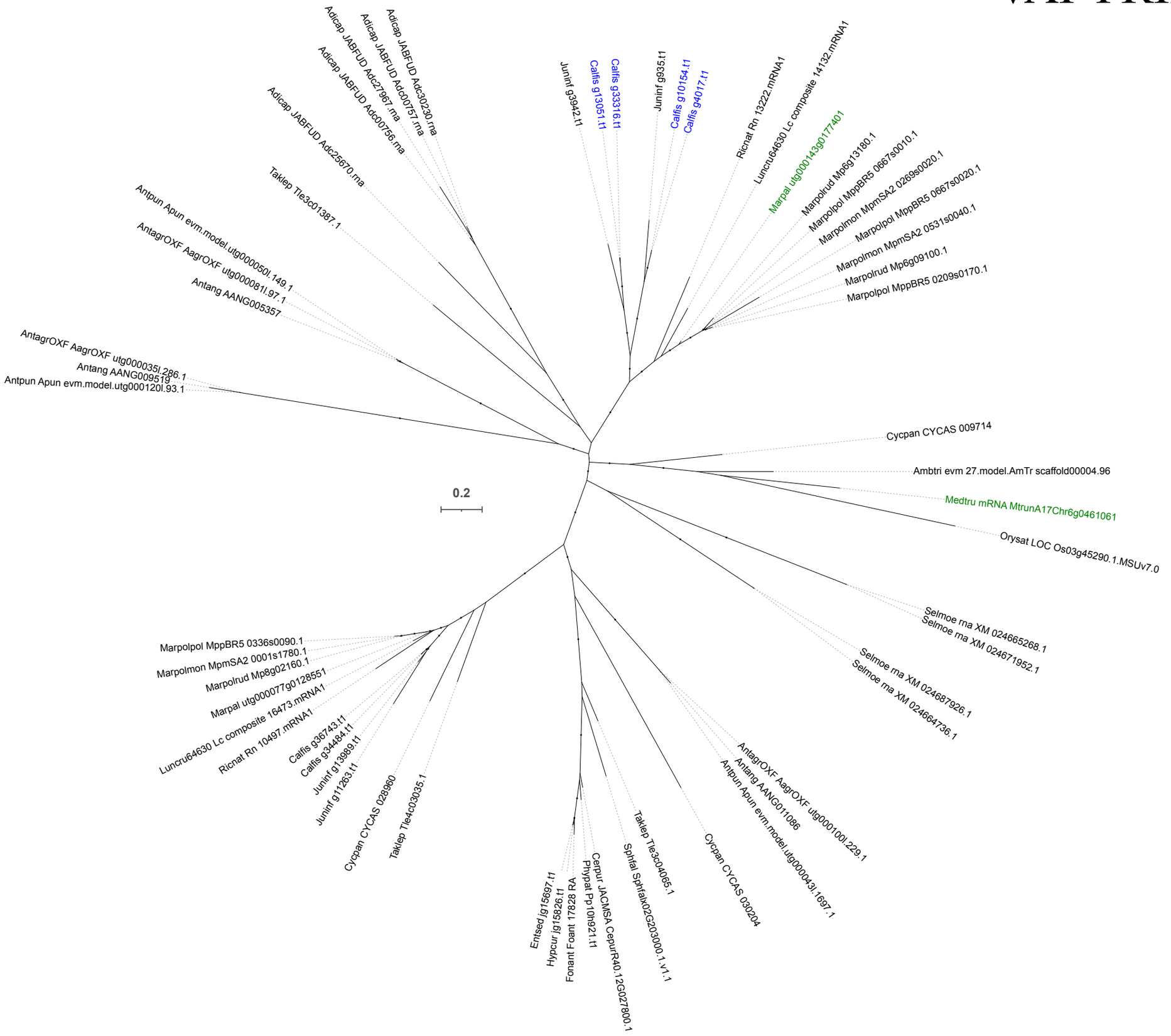
# SymRK


Orysat LOC Os07g38070.1.MSUv7.0
Ambtri evm 27.model.AmTr scaffold00046.162
Medtru mRNA MtrunA17Chr5g0411631
Cycpan CYCAS 004183
Adicap JABFUD Adc03126.rna
Selmoe rna XM 024666164.1
Selmoe rna XM 024663931.1
Selmoe rna XM 002979516.2
Selmoe rna XM 002977662.2
0.2
Taklep Tle3c05471.1
Calfis g6484.t1
Calfis g19379.t1
Juninf g8903.t1
Antang AANG006003
AntagrOXF AagrOXF utg000126l.231.8
Antpun Apun evm.model.utg000045l.172.1
Marpal utg000051g0090241
Luncru64630 Lc composite 20805.mRNA1

SYN


Ricnat Rn 09111.mRNA1
Ricnat Rn 06715.mRNA1
Ricnat Rn 07289.mRNA1
Marpal utg000126g0168881
Marpolmon MpmSA2 0021s1040.1
Marpolpol MppBR5 0030s0400.1
Marpolrud Mp2g19600.1
Antang AANG007961
Antang AANG002887
AntagrOXF AagrOXF utg000033l.82.1
Antpun Apun evm.model.utg000041l.94.1
Sphfal Sphfalx03G082000.1.v1.1
Sphfal Sphfalx13G070700.1.v1.1
Sphfal Sphfalx05G038500.1.v1.1
Taklep Tle2c07189.1
Entsed jg920.t1
Hypcur jg24071.t1
Entsed jg4126.t1
Phypat Pp04h1206.t1
Phypat Pp03h500.t1
Cerpur JACMSA CepurR40.1G164500.1
Fonant Foant 06887 RA
Entsed jg10528.t1
Hypcur jg12907.t1
Isosin Chr4.3433
Isosin Chr3.3502
Selmoe rna XM 002990714.2
Selmoe rna XM 002986607.2
Selmoe rna XM 024681374.1
Selmoe rna XM 002987109.2
Isosin Chr4.2627
Isosin Chr3.2686.2
Azofil Azfi s0113.g045884
Adicap JABFUD Adc30934.ma
Adicap JABFUD Adc03430.ma
Azofil Azfi s0059.g034575
Adicap JABFUD Adc16061.ma
Cycpan CYCAS 007786
Medtru mRNA MtrunA17Chr2g0322801
Ambtri evm.27.model.AmTr scaffold00004.233
Ambtri evm.27.model.AmTr scaffold00004.123
Orysat LOC Os06g07200.1.MSUv7.0
Orysat LOC Os07g07000.1.MSUv7.0
Aratha AT5G08080.3.TAIR10
Aratha AT3G03800.1.TAIR10
Medtru mRNA MtrunA17Chr6g0461261
Ricnat Rn 01054.mRNA1
Marpal utg000074g0124311
Marpolpol MppBR5 0076s0390.1
Marpolrud Mp7g13190.1
Marpolmon MpmSA2 0077s0010.1
Marpolrud Mp7g04280.1
Marpolmon MpmSA2 0202s0410.1
Luncru64630 Lc composite 00194.mRNA1
Juninf 9378.t1
Calfis 931465.t1
Calfis 931971.t1
0.2

# VAPYRIN


Adicap JABFUD Adc30230.ma
Adicap JABFUD Adc00757.ma
Adicap JABFUD Adc27967.ma
Adicap JABFUD Adc00756.ma
Adicap JABFUD Adc25670.ma
Taklep Tle3c01387.1
Antpun Apun evm.model.utg000050l.149.1
AntagrOXF AagrOXF utg000081l.97.1
Antang AANG005357
AntagrOXF AagrOXF utg000035l.286.1
Antang AANG009519
Antpun Apun evm.model.utg000120l.93.1
Juninf g3942.t1
Calfis g33316.t1
Calfis g13051.t1
Juninf g935.t1
Calfis g10154.t1
Calfis g4017.t1
Ricnat Rn 13222.mRNA1
Luncru64630 Lc composite 14132.mRNA1
Marpal utg000143g0177401
Marpolrud Mp6g13180.1
Marpolpol MppBR5 0667s0010.1
Marpolmon MpmSA2 0269s0020.1
Marpolpol MppBR5 0667s0020.1
Marpolmon MpmSA2 0531s0040.1
Marpolrud Mp6g09100.1
Marpolpol MppBR5 0209s0170.1
Cycpan CYCAS 009714
Ambtri evm 27.model.AmTr scaffold00004.96
Medtru mRNA MtrunA17Chr6g0461061
Orysat LOC Os03g45290.1.MSUv7.0
Selmoe rna XM 024665268.1
Selmoe rna XM 024671952.1
Selmoe rna XM 024687926.1
Selmoe rna XM 024664736.1
AntagrOXF AagrOXF utg000100l.229.1
Antang AANG011086
Antpun Apun evm.model.utg000043l.1697.1
Cycpan CYCAS 030204
Taklep Tle3c04065.1
Sphfal Sphfalx02G203000.1.v1.1
Cepur JACMSA CepurR40.12G027800.1
Phypat Pp1h921.t1
Fonant Foant 17828 RA
Hypcur jg15826.t1
Entsed jg15697.t1
Marpolpol MppBR5 0336s0090.1
Marpolmon MpmSA2 0001s1780.1
Marpolrud Mp8g02160.1
Marpal utg000077g0128551
Luncru64630 Lc composite 16473.mRNA1
Ricnat Rn 10497.mRNA1
Calfis g36743.t1
Calfis g34484.t1
Juninf g13989.t1
Juninf g11263.t1
Cycpan CYCAS 028960
Taklep Tle4c03035.1
0.2

# Tables

**Table 1. Genome assembly statistics for the leafy liverworts *Solenostoma infuscum* and *Calypogeia fissa*.**

| | | *Solenostoma infuscum* | *Calypogeia fissa* |
|---|---|---|---|
| **Assembly size (bp)** | | 383,453,254 | 763,616,257 |
| **Scaffolds** | | 4,616 | 82 |
| **N50 length (bp)** | | 244,190 | 15,851,540 |
| **BUSCO genome completion (%)** | **Complete** | 95.8 | 96 |
| | **Single Copy** | 92.5 | 2.1 |
| | **Duplicated** | 3.3 | 93.9 |
| | **Fragmented** | 0.5 | 0.5 |
| | **Missing** | 3.8 | 3.5 |
| **Coding sequences** | | 15,468 | 38,150 |