



**HAL**  
open science

# The genome of *Geosiphon pyriformis* reveals ancestral traits linked to the emergence of the arbuscular mycorrhizal symbiosis

Mathu Malar C, Manuela Krüger, Claudia Krüger, Yan Wang, Jason E Stajich, Jean Keller, Eric C H Chen, Gokalp Yildirim, Matthew Villeneuve-Laroche, Christophe Roux, et al.

## ► To cite this version:

Mathu Malar C, Manuela Krüger, Claudia Krüger, Yan Wang, Jason E Stajich, et al.. The genome of *Geosiphon pyriformis* reveals ancestral traits linked to the emergence of the arbuscular mycorrhizal symbiosis. *Current Biology - CB*, 2021, 31 (7), pp.1570-1577.e4. 10.1016/j.cub.2021.01.058 . hal-03327905

**HAL Id: hal-03327905**

**<https://hal.science/hal-03327905>**

Submitted on 1 Sep 2021

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

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

1 **The genome of *Geosiphon pyriformis* reveals ancestral traits linked to the**  
2 **emergence of the arbuscular mycorrhizal symbiosis**

3 Mathu Malar C <sup>1\*</sup>, Manuela Krüger <sup>2,3,\*</sup>, Claudia Krüger <sup>2,3\*</sup>, Yan Wang <sup>4,5</sup>, Jason E. Stajich <sup>6</sup>,  
4 Jean Keller <sup>7</sup>, Eric C.H. Chen<sup>1</sup>, Gokalp Yildirim <sup>1</sup>, Matthew Villeneuve-Laroche<sup>1</sup>, Christophe  
5 Roux <sup>7</sup>, Pierre-Marc Delaux <sup>7</sup> and Nicolas Corradi <sup>1#</sup>

6 <sup>1</sup> Department of Biology, University of Ottawa. Ottawa, Canada.

7 <sup>2</sup> Institute of Botany, The Czech Academy of Sciences. Průhonice, Czech Republic

8 <sup>3</sup> Institute of Experimental Botany, The Czech Academy of Science, Prague, Czech Republic

9 <sup>4</sup> Department of Biological Sciences, University of Toronto Scarborough, Toronto, Canada

10 <sup>5</sup> Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada

11 <sup>6</sup> Department of Microbiology & Plant Pathology and Institute for Integrative Genome Biology,  
12 University of California–Riverside, Riverside, California. USA

13 <sup>7</sup> Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, UPS, CNRS 24  
14 Chemin de Borde Rouge-Auzeville, Castanet-Tolosan, France

15 \*contributed equally

16 #Lead contact

17 Corresponding authors

18 Nicolas Corradi email: [ncorradi@uottawa.ca](mailto:ncorradi@uottawa.ca)

19 Mathu Malar email: [mmadhubioinfo@gmail.com](mailto:mmadhubioinfo@gmail.com)

20 Nicolas Corradi Twitter: @Blunt\_Microbe

21 Mathu Malar Twitter: @madhubioinfo

## 1 **Summary**

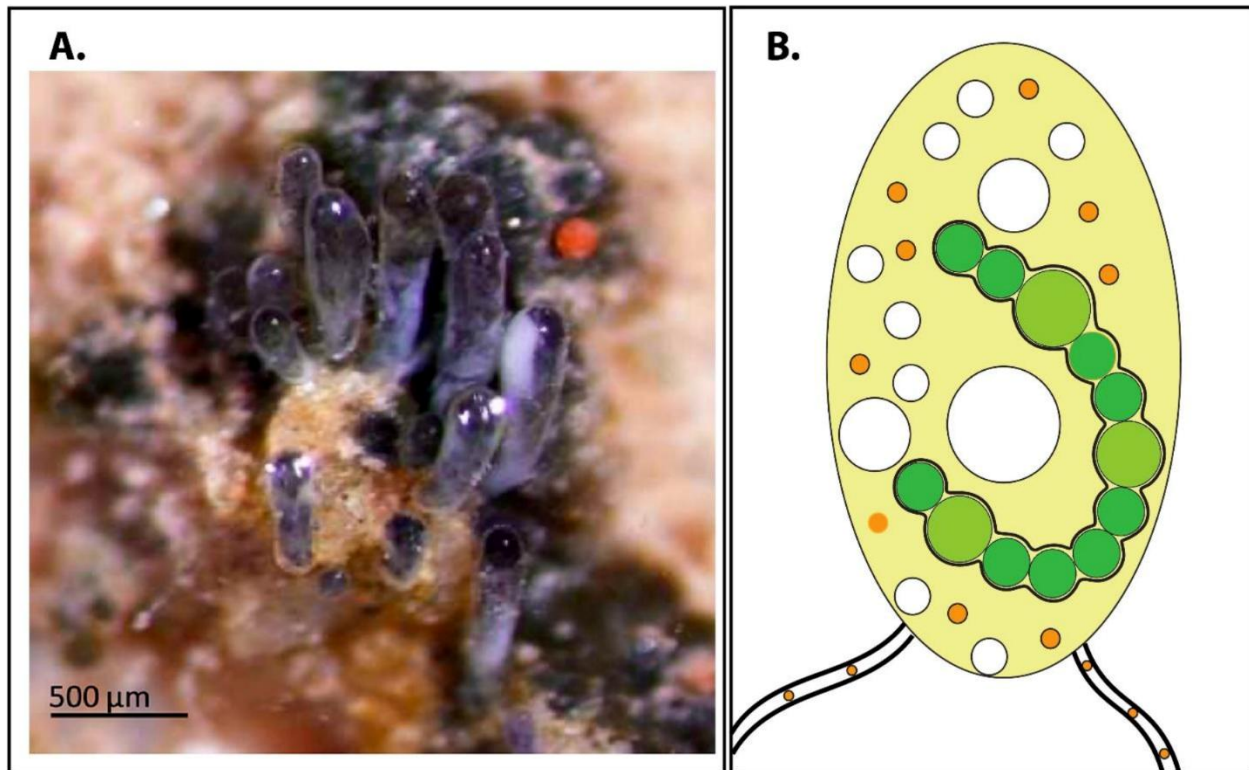
2 Arbuscular mycorrhizal fungi (AMF) (subphylum Glomeromycotina) [1] are amongst the most  
3 prominent symbionts and form the Arbuscular Mycorrhizal symbiosis (AMS) with over 70% of  
4 known land plants [2] [3]. AMS allows plants to efficiently acquire poorly soluble soil nutrients  
5 [4] and AMF to receive photosynthetically fixed carbohydrates. This plant-fungus symbiosis  
6 dates back more than 400 million years [5], and is thought to be one of the key innovations that  
7 allowed the colonization of lands by plants [6] . Genomic and genetic analyses in diverse plant  
8 species started to reveal the molecular mechanisms that allowed the evolution of this symbiosis  
9 on the host side, but how and when AMS abilities emerged in AMF remains elusive.  
10 Comparative phylogenomics could be used to understand the evolution of AMS [7,8]. However,  
11 the availability of genome data covering basal AMF phylogenetic nodes (Archaeosporales,  
12 Paraglomerales) is presently based on fragmentary protein coding datasets [9]. *Geosiphon*  
13 *pyriformis* (Archaeosporales) is the only fungus known to produce endosymbiosis with nitrogen-  
14 fixing cyanobacteria (*Nostoc punctiforme*) presumably representing the ancestral AMF state [10–  
15 12]. Unlike other AMF, it forms long fungal cells (‘bladders’) that enclose cyanobacteria. Once  
16 in the bladder, the cyanobacteria are photosynthetically active and fix nitrogen, receiving  
17 inorganic nutrients and water from the fungus. Arguably, *G. pyriformis* represents an ideal  
18 candidate to investigate the origin of AMS and the emergence of a unique endosymbiosis. Here,  
19 we aimed to advance knowledge in these questions by sequencing the genome of *G. pyriformis*,  
20 using a re-discovered isolate.

21 **Keywords:** *Geosiphon pyriformis*, homokaryon, bladders, binning, metagenome, symbiosis,  
22 cyanobacteria, *Nostoc Punctiforme*

## 23 **Results**

### 24 *General genome characteristics of Geosiphon pyriformis*

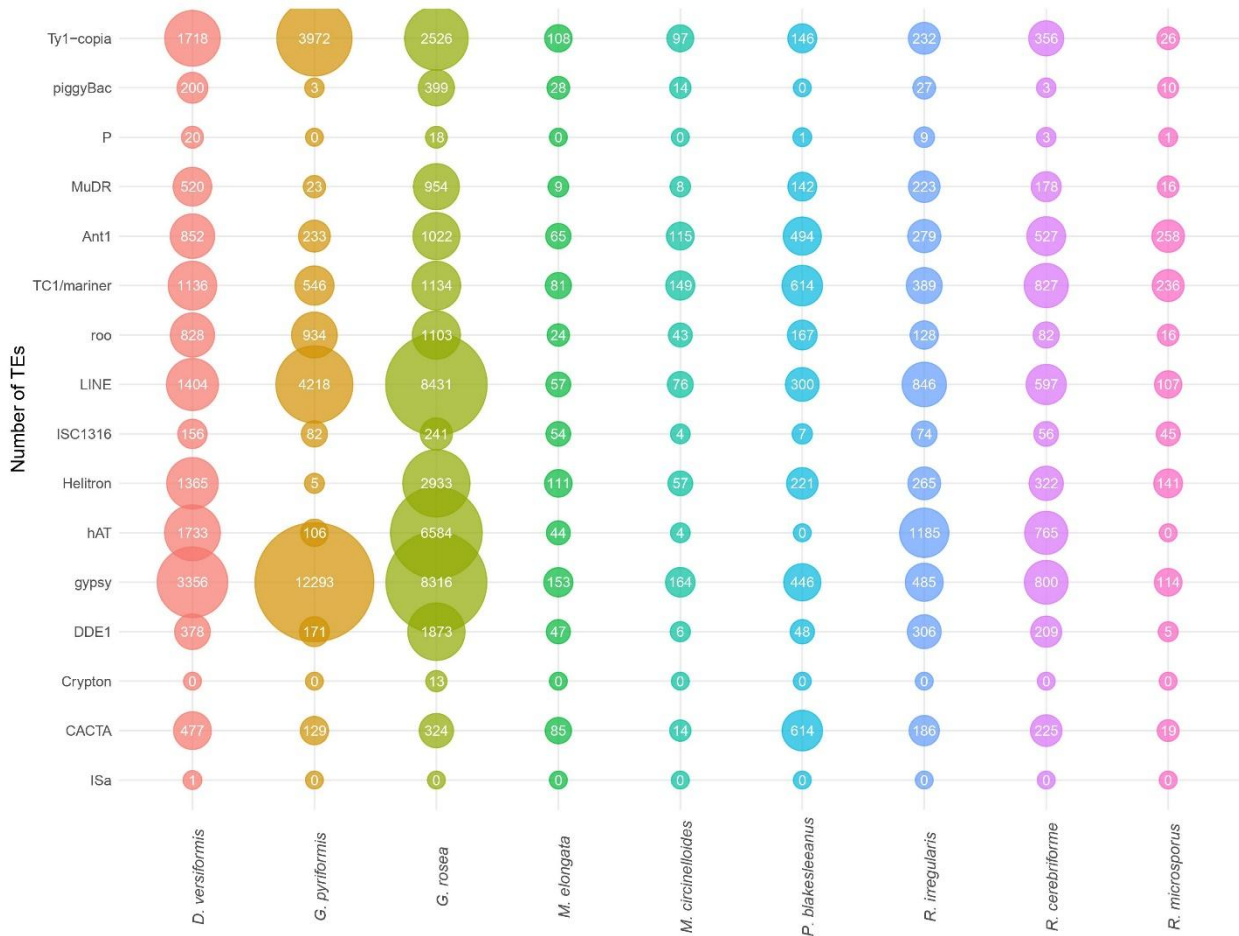
25 The only known culture of *G. pyriformis* was lost over a decade ago (A. Schüßler, pers. comm).  
26 In an attempt to rediscover *G. pyriformis*, we searched and identified symbiotic bladders of the  
27 *G. pyriformis*-*N. punctiforme* symbiosis (Figure 1) at the only known stable habitat of this  
28 species in the Spessart Mountains near the village of Bieber in Germany [13].



1  
2 **Figure 1: a.** Image of *G. pyriformis* bladders in soil from its natural habitat in the Spessart  
3 mountains. **b.** Schematic representation of *G. pyriformis* bladders containing *Nostoc* cells (based  
4 on M. Kluge 2002). *Nostoc* cells are shown in dark green, and heterocysts (differentiated cell that  
5 carries out nitrogen fixation) are shown in light green. Bladders contain *G. pyriformis* nuclei  
6 (orange), and several vacuoles (white). Aseptate hyphae reach out and extend from the bladders.  
7 Upon cultivation, we extracted DNA and RNA from active bladders. Total DNA was subjected  
8 to 5Kb-mate-pair and 125 bp paired-end Illumina sequencing, producing respectively 47 and 81  
9 million 125 bp paired ends and 5kb mate-pairs reads. *G. pyriformis* reads were identified using a  
10 read binning approach recently implemented to assemble the genome of *Diversispora epigaea*  
11 *epigaea* [14], and upon identification these were assembled into a 129 MB assembly and 795  
12 scaffolds with an average read coverage of 118X. In parallel, total RNA was subjected to 150 bp  
13 paired-end Illumina sequencing. The resulting RNA-seq reads were mapped onto the *G.*  
14 *pyriformis* genome assembly using STAR [15] and used for genome annotation after  
15 implementing RepeatMasker [16]. This procedure identified of 24195 genes in *G. pyriformis*,  
16 resulting in a BUSCO gene repertoire completeness of 96.2% (3.1% complete duplicated). The  
17 gene counts, estimated genome size, and genome statistics are all similar to those of model AMF

1 species [17,18] and are indicative of high genome completeness (Table 1, see figure of “K-mer  
 2 distribution of filtered illumina genomic reads of *G. pyriformis*” in Data S1 and see table of  
 3 “Genome completeness using Busco” in Data S1). SignalP showed that, among all the genes  
 4 identified in *G. pyriformis*, 365 represent putative secreted proteins (see table of “Secretome  
 5 proteins from genomes” in Data S1) and 27% of these are candidate effectors (see table of  
 6 “Effector prediction on *G. Pyriformis* genome “in Data S1). We also identified 19 putative  
 7 secreted CAZymes in *G. pyriformis*; in line with numbers found in other AMF species (Table S1  
 8 and Table S2).

9



10

11 **Figure 2:** Bubble plot containing all transposable elements in the genomes of glomeromycotina  
 12 and mucoromycotina genomes used in this study. The figure shows the expansion of Gypsy  
 13 elements in *G. pyriformis*.

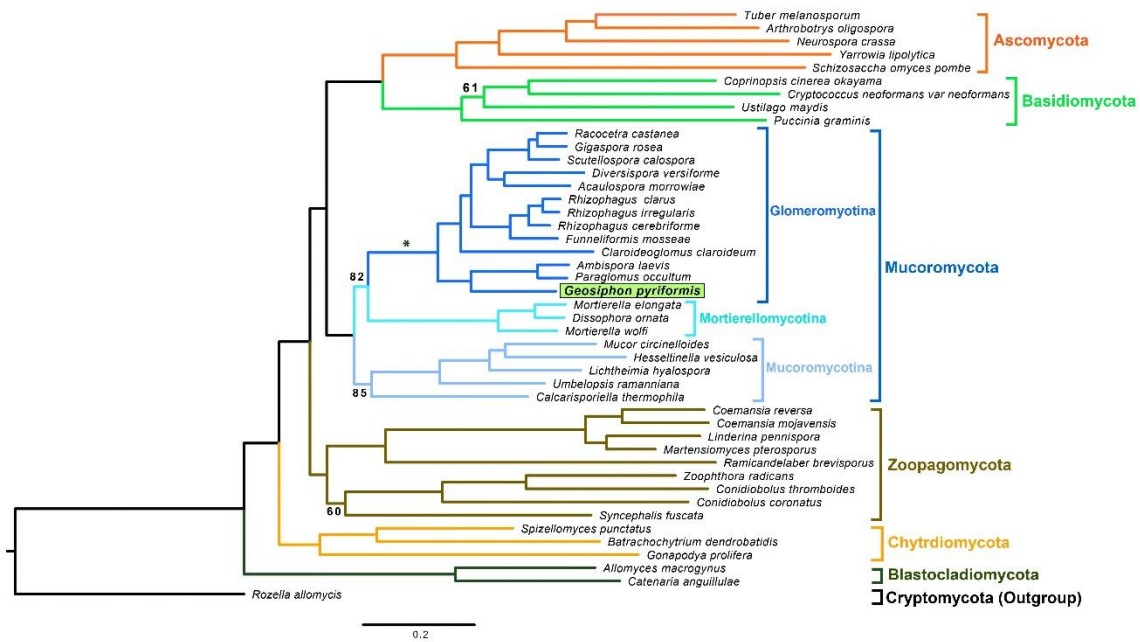
1  
2 AMF genomes carry a substantial fraction of transposable elements (TE) [18–21], and we found  
3 that *G. pyriformis* has undergone similar TE expansions. The expansion of Gypsy transposable  
4 elements in *G. pyriformis* is evident in comparison to all other AMF genomes (Figure 2). With  
5 regards to TE, we find no evidence that *G. pyriformis* (see figure of “Heatmap showing evidence  
6 for a “one-speed genome” in *Geosiphon Pyriformis*” in Data S1) carries a two-speed genome  
7 [22]. Two-speed genomes are characterized by the presence of TE-poor and gene dense regions  
8 that a clearly separate from others that contain rapidly evolving genomic regions that usually  
9 carry less genes, abundant TE, and other repeat elements [22,23].

10 Using genome data and single nucleus data it was recently shown that AMF carry two genome  
11 organizations – i.e., homokaryotic (co-existing nuclei carry one parental haploid genotype) or  
12 dikaryotic (two parental genotypes co-exist in the mycelium) [21,24–26]. Mapping reads onto  
13 the *G. pyriformis* genome revealed reduced levels of polymorphism (0.5 SNP/Kb) and allelic  
14 frequencies suggesting that this species is carries low a low nuclear diversity and likely  
15 homokaryotic (see figure of “Genome wide allele frequency of *G. Pyriformis* shows it is a  
16 homokaryon” in Data S1).

### 17 ***Placement of G. pyriformis based on phylogenomics***

18 The *G. pyriformis* genome annotation was used to identify the phylogenetic placement of this  
19 species using amino-acid sequences. In this case, we used a set of 434 conserved fungal single  
20 copy genes (data available at DOI: 10.5281/zenodo.1413687) to construct a phylogenetic tree of  
21 the fungal kingdom. Phylogenomics supports the monophyly of Glomeromycotina and its close  
22 relationship with Mortierellomycotina within the phylum Mucoromycota [1] (Figure 3). Within  
23 Glomeromycotina, *G. pyriformis* groups within a monophyletic clade with *Ambispora leptoticha*  
24 and *Paraglomus occultum*, which diverged around 287 MYA (Figure S1). This clade is distinct  
25 from more diverged nodes that contain sequenced representatives from Glomerales and  
26 Diversisporales [27]. The current placement of *G. pyriformis* as a sister lineage to *A. leptoticha*  
27 and *P. occultum* has full statistical support, and is favored by 73% of the gene sequences we  
28 used. Alternative topologies that place, for example, *G. pyriformis* at a basal node to all AMF  
29 (Alt-T1; Figure S2, see table of “Results of alternative tree topology test for phylogenetic tree “ in  
30 Data S1) or as being associated with Glomerales or Diversisporales (Alt-T2 and Alt-T3; Figure

1 S2, see table of “Results of alternative tree topology test for phylogenetic tree “ in Data S1) were  
 2 all rejected significantly using statistical tests implemented in IQ-TREE, including the KH, SH,  
 3 ELW, and AU tests (see table of “Results of alternative tree topology test for phylogenetic tree “  
 4 in Data S1).



5  
 6 **Figure 3:** Phylogenetic tree representing the evolutionary relationships of fungi and placement  
 7 of *G. pyriformis* in Glomeromycotina clade. The tree was resolved using maximum likelihood  
 8 phylogenetic reconstruction with IQ-TREE on a concatenated alignment of 434 protein coding  
 9 genes. Numbers indicate nodes with less than 100% bootstrap support. Branches are coloured  
 10 according to their phylum.

11 Phylogenetic tests for all alternative topological placements of *G. pyriformis* were rejected. The  
 12 asterisk denotes the location where inferred gene losses and gains occurred in the MRCA of  
 13 Glomeromycotina.

14 ***The genome of Geosiphon pyriformis uncovers shared gene features in the Glomeromycotina***

15 Phylogenomics revealed that *G. pyriformis* is a member of a clade that diverged early from the  
 16 lineage encompassing the already sampled Diversisporales and Glomerales. As such, the *G.*  
 17 *pyriformis* genome fills the gap in the genomic coverage of major AMF phylogenetic clades.  
 18 With this data in hand, we first searched for genetic features that arose in the most recent

1 common ancestors (MRCA) of all Glomeromycotina by comparing orthogroups from five  
2 available AMF genomes and four other members of the Mucoromycota as outgroups using  
3 OrthoFinder [28]. This analysis identified 661 gains and 344 losses that occurred before the  
4 divergence of the Glomeromycotina (Table S3 and see table of “Orthogroups lost in the MRCA  
5 of Glomeromycotina” in Data S1).

6 Among the 344 orthogroups classified as lost in the MRCA of the Glomeromycotina, we note the  
7 missing key enzymes involved in essential metabolic functions, such as sugar and thiamine  
8 metabolisms, or in the biosynthesis of fatty acids. These genes are also referred to as “Missing  
9 Glomeromycotina Core Genes” (MGCGs; see table of “Orthogroups lost in the MRCA of  
10 Glomeromycotina” in Data S1, see table of “Missing MRCA genes across glomeromycotina core  
11 genes” in Data S1), and our analysis reveals that these have also been lost by *G. pyriformis*.  
12 Other key losses that affect all sequenced Glomeromycotina include enzymes that actively  
13 degrade plant cell wall (see table of “Presence and absence of plant and fungal cell wall related  
14 CAZymes” in Data S1) . Among the 661 orthogroups gained in the MRCA, most encode for  
15 proteins involved in signaling pathways (e.g., protein kinases), protein–protein interactions (e.g.,  
16 the tetratricopeptide repeat, Sell1, the homodimerization BTB (Broad- Complex, Tramtrack and  
17 Bric a brac), and WD- 40 domain- containing proteins) and High Mobility Box (HMG) (see  
18 table of “Functions of genes which are gained in orthogroups” in Data S1).

19 Comparative genomics also showed that *G. pyriformis* carries the same signatures of sexual  
20 reproduction found in AMF relatives. These include a complete set of meiosis-specific genes  
21 (see table of “Identification of Meiosis specific genes” in Data S1) [29,30], and a highly  
22 conserved genomic locus with architecture and sequence similarity to the mating-type (MAT)  
23 locus of basidiomycetes [21,31,32] (see figure of “Transcriptional directions of the putative  
24 AMF mating-type locus” in Data S1).

25

26

### 27 ***Regulation of orthogroups gained in the MRCA of Glomeromycotina***

28 We investigated available gene expression data from the model AMF *Rhizophagus irregularis*,  
29 and found that 7 and 39 orthogroups (with a total of 8 and 272 genes; see table of “Upregulation



1 of ortholog genes from Rhiir2 genome “in Data S1 and see table of “Downregulation of  
2 Orthogroups genes from Rhiir2” in Data S1) gained in the MRCA of *Glomeromycotina* are  
3 respectively upregulated and downregulated across all four experimental conditions in the model  
4 AMF *Rhizophagus irregularis*.

5 These conditions include symbiotic associations between *R. irregularis* and distinct plant hosts  
6 (*Medicago truncatula*, *Brachypodium distachyon*) [33–35], and others based on laser-capture  
7 microdissection arbuscule-specific gene expression [33,34,36]. We investigated the putative  
8 function of these differentially regulated genes by identifying protein motifs along their coding  
9 sequences, and found that these are involved in a myriad putative cellular function, though these  
10 mostly include protein tyrosine kinases, cytochrome p450, as well as FAD binding (see table of  
11 “Upregulation of ortholog genes from Rhiir2 genome” and “Downregulation of Orthogroups  
12 genes from Rhiir2” in Data S1). Lastly, one differentially regulated OG (OG0001728) shows  
13 evidence of originating from horizontal gene transfer from bacteria – i.e., this orthogroup is  
14 shared between bacteria and fungi (Figure S3).

#### 15 ***Distinct genomic features of Geosiphon pyriformis***

16 Besides the evolution of AMS, the acquisition of the *G. pyriformis* genome offers an opportunity  
17 to identify innovations linked to the emergence of the only known cyanobacteria – fungus  
18 endosymbiosis. To identify such innovations, hierarchical clustering and abundance of Pfam  
19 domains was performed using available genomes in the Glomeromycotina and representatives of  
20 Mucoromycotina and Mortierellomycotina (Figure S4). This analysis revealed a significant  
21 overrepresentation of 16 protein domains in *G. pyriformis* compared to relatives in the  
22 Mucoromycota – e.g., Lipase\_3, RNase\_H, Retrotrans gag domains, dUTPase, Spuma\_A9PTase,  
23 Myb\_DNA-bind\_6 (see table of “Pfam domain counts from genomes of mucoromycota used in  
24 this study” and “p values calculated by fisher's test” in Data S1).

25 We also sought evidence of horizontal gene transfers (HGT) between partners of the unique  
26 *Geosiphon-Nostoc* endosymbiosis, and found 18 genes with potential bacterial within in the *G.*  
27 *pyriformis* genome (see table of “HGT containing genes with pfam domains” in Data S1).  
28 Among putative HGT, two are protein encoding genes with significant sequence conservation  
29 with *Nostoc* and Gamma proteobacteria homologues (see figure of “Phylogenetic tree showing  
30 evidence of Horizontal gene transfer of Selenium binding protein” and “Phylogenetic tree

1 showing evidence of horizontal gene transfer in Molybdenum cofactor carrier” in Data S1). All  
2 putative HGT are located within contigs with average coverage and surrounded by genes of  
3 AMF origin, suggesting these do not represent contaminants.

#### 4 **Discussion**

##### 5 ***MRCA of all extant Glomeromycotina carried hallmarks of mutualism and obligate biotrophy***

6 Genome data from a representative of the basal node of the AMF phylogeny filled an important  
7 gap in understanding the origin of AMS. Specifically, it allowed us to conclude that the MRCA  
8 of all extant Glomeromycotina carried hallmarks of mutualism and obligate biotrophy – i.e., a  
9 lack of genes for fatty acids and thiamine biosynthesis and nutrition, and a reduced number of  
10 genes that actively degrade plant cells. As such, the mechanisms involved in AMS appeared  
11 prior to the emergence of Glomeromycotina and a represent a synapomorphy of this sub-phylum.  
12 *G. pyriformis* has also conserved genomic signatures of sexual reproduction, as well as an  
13 apparent low nuclear polymorphism. Both traits are thus conserved across Glomeromycotina and  
14 are in stark contrast with the notion that these organisms represent an ancient asexual lineage.

15 The retention of a conserved Glomeromycotina gene set in *G. pyriformis*, including a sub-set of  
16 these involved in plant cell wall degradation, indicates that this species might be possible to form  
17 mycorrhizae, although this has never been observed. As such, this retention could reflect an  
18 intrinsic capacity for *G. pyriformis* to undergo classic (but rare) mycorrhizal associations with  
19 plants under the right conditions. Although speculative at this point, this hypothesis is supported  
20 by the identification of rare *Geosiphon*-like sequences in environmental samples [37–39].

##### 21 ***Novel symbiotic abilities and horizontal gene transfers in G. pyriformis***

22 As a result of losses in fatty acid biosynthesis genes, AMF are entirely dependent on the host  
23 plant they associate with to obtain lipids [40–43]. Within this context, our findings suggest that,  
24 during the switch from regular AMS to a fungal- cyanobacteria symbiosis, *G. pyriformis* has  
25 evolved novel strategies to obtain lipids from its new host through the expansion of specific gene  
26 motifs. Specifically, the *G. pyriformis* genome carries a striking over-representation of Lipase 3  
27 protein domains that hydrolyze ester linkages of fatty acids. As *Nostoc* spp is known to produce  
28 a wide variety of extracellular lipids in high amounts [40–43], it is possible that these abundant  
29 lipids are released in the environment (like many other cellular compounds released by

1 cyanobacteria [44–47]) and are then broken down by lipases to be used as an energy resource by  
2 *G. pyriformis*.

3 As we find evidence of bacteria-like genes in the *G. pyriformis* genome, our work also suggests  
4 that the co-existence of multiple endosymbionts and *G. pyriformis* nuclei within restricted  
5 bladders offers some opportunities for horizontal gene exchange. Although none of the putative  
6 bacterial genes we identified in *G. pyriformis* are functionally related, there is evidence that one  
7 is differentially regulated during AMF symbiosis, supporting the notion that bacteria-like genes  
8 can play a major role in fungal evolution [14,48].

### 9 ***Identification of a putative AMF core-symbiotic toolkit***

10 The *G. pyriformis* genome also enabled the identification of a putative core AMF symbiotic  
11 toolkit conserved in all the sampled Glomeromycotina. This set of genes is differentially  
12 regulated in model AMF during symbiotic interactions with different plant hosts, including  
13 dicots, monocots and non-vascular plants, and thus provides a basis for future research on  
14 symbiosis-related mechanisms in these plant symbionts. The identification of a core set of gene  
15 gains specifically regulated during mycorrhizal symbiosis, and their conservation across the  
16 Glomeromycotina phylogeny, also provides support for the early emergence of symbiosis-  
17 specific gene functions in AMF over 400 million years ago, contemporaneously with the  
18 evolution of the first land plants [49,50]. Lastly as genetic transformation is currently unfeasible  
19 in Glomeromycotina, only assumptions can be proposed for the function of these putative core  
20 genes. However, as some encode for chitin synthases, one attractive hypothesis could be that  
21 some evolved for the production of short-chain chitooligosaccharides or lipo-  
22 chitoligosaccharides that are known symbiotic signals triggering the activation of the symbiotic  
23 program on the host plant [2,51,52].

### 24 **Acknowledgements**

25 We thank Vasilis Kokkoris and Allison MacLean for comments on an earlier version of the  
26 manuscript. N.C research is funded by the discovery program of the Natural Sciences and  
27 Engineering Research council (RGPIN-2020-05643) and the Discovery Accelerator Supplements  
28 program (RGPAS-2020-00033). N.C is a University of Ottawa Research chair in Microbial  
29 Genomics. M.K. and C.K. were funded by the Czech Sciences Foundation (GAČR) as Junior  
30 grant with the project number GJ16-16406Y. This work was also supported by the Agence

1 Nationale de la Recherche (ANR) grant EVOLSYM (ANR-17-CE20-0006-01) to P.M.D, by the  
2 Bill and Melinda Gates Foundation as Engineering the Nitrogen Symbiosis for Africa  
3 (OPP1172165) to PM. D and by Laboratoire de Recherche en Sciences Végétales (LRSV)  
4 laboratory, which belongs to the TULIP Laboratoire d'Excellence (ANR-10-LABX-41). J.E.S. is  
5 a CIFAR Fellow in the program Fungal Kingdom: Threats and Opportunities. Y.W. and J.E.S.  
6 were supported by US National Science Foundation grants DEB-1441715 and DEB-1557110.

### 7 **Author Contributions**

8 MMC, MK, CK, YW, JS, PM, and NC planned and designed the research, wrote the Manuscript  
9 and helped with data analysis. MMC carried out genome annotation and bioinformatics analysis.  
10 YW performed phylogenetic and molecular dating analysis. ECHC and GY produced repeat  
11 analysis with MMC. JK performed ortholog analysis. CR performed the transcriptome analysis.  
12 MVL helped improving the quality of images. MK and CK produced biological materials. NC  
13 supervised all process.

### 14 **Declaration of Interests**

15 The authors declare no competing interests.

### 16 **Star Methods**

### 17 **Resource availability**

18 Further information and requests for resources and reagents should be directed to and will be  
19 fulfilled by the Lead Contact, Nicolas Corradi ([ncorradi@uottawa.ca](mailto:ncorradi@uottawa.ca)).

### 20 **Materials availability**

21 This study did not generate new unique reagents.

### 22 **Experimental Model and Subject Details**

23 AM Fungi *G. pyriformis* were used for this study. Genome assembly is submitted in NCBI  
24 accession number of AAOMT000000000.

### 25 **Data and Code Availability**

26 Genome assembly is available in NCBI with accession number of JAAOMT000000000. Genome  
27 sequencing reads are submitted in SRA with accession number of SRR11466073, Bioproject

1 PRJNA610605, Biosample SAMN14307302. RNA-seq reads are available in SRA with  
2 accession of SRR12018969, SRR12018968, SRR12018970. All scripts used to analyze are  
3 archived in <https://github.com/madhubioinfo/Geosiphon>.

4

## 5 **METHOD DETAILS**

### 6 **Cultivation of *G. pyriformis* samples from natural habitat**

7 *G. pyriformis* was sampled during autumn in the only known stable habitat near the village  
8 Bieber in the Spessart (Germany). Active bladders of the *Geosiphon-Nostoc* endosymbiosis were  
9 found in slightly acidic soil (pH 5). The bladders occurred close to the hornwort *Anthoceros* spp.  
10 and the liverwort *Blasia pusilla* L., as these plants harbor the cyanobacteria needed to trigger the  
11 *Geosiphon-Nostoc* endosymbiosis. After sampling spores and bladders were transferred to the  
12 institute in Průhonice and cultured in beakers [53,54], which contain a small pot with a sterile  
13 mixture of sand and soil (from the original habitat). The cultures were grown in a climate  
14 chamber at 18°C with 14 h light and 10 h night. The substrate is kept wet by a filter paper, which  
15 reaches from the substrate into a water reservoir in the beaker. To be maintained over time,  
16 cyanobacteria be frequently added to the cultures. For our cultures, *Nostoc punctiformis* was  
17 obtained from the Culture Collection of Algae (SAG) at the University of Göttingen (Germany)  
18 as strain SAG69.79 [53,54].

### 19 ***Genome and transcriptome sequencing and assembly***

20 High quality DNA was extracted from active bladders of *G. pyriformis* and *Nostoc punctiforme*  
21 using the NucleoSpinII Plant kit (Machery-Nagel) and purified with the genomic DNA clean-up  
22 kit (Machery-Nagel) using the manufactures recommendations. Total DNA was sent to Fasteris  
23 (Switzerland) for library Illumina library preparation and sequencing using on 150 paired end  
24 and 5kb mate pairs inserts (illumina Nextera mate pair kit). Sequencing was performed using the  
25 Illumina Hiseq 4000 platform. Total RNA was extracted using the RNeasy-Mini Kit (Qiagen) as  
26 per instructions of the manufacturer for library RNA-seq Illumina library preparation with  
27 sequencing of 150 cycles and paired ends.

28 Poor quality and adapter sequences were trimmed using Trimmomatic [55] with parameters the  
29 following parameters of ILLUMINACLIP:2:30:10 SLIDINGWINDOW:5:20 LEADING:5

1 TRAILING:5 MINLEN:50. The resulting 1 GB of non-redundant metagenome reads were  
2 assembled using metaSPAdes V3.12.0 [56]. Assembled contigs were binned on the basis of tetra  
3 nucleotide signature using CONCOCT [57], following part of the procedure used to assembly the  
4 genome of *Diversispora epigaea* [14]. Binned clusters were annotated using BLAST v 2.6.0+  
5 [58,59] and clusters containing bacterial hits were removed. Using this approach, 21 bona-fide  
6 AMF clusters were retained, and used as reference to filter original paired-end and mate pair  
7 reads with BLAT v. 36x1 [60]. The reads which had mapped to the filtered contigs and matched  
8 by BLAT were then extracted to build cleaned sequence libraries that were assembled with  
9 MaSuRCA 3.3.0 [61]. Additional round of nr BLAST searches on MaSuRCA assembled contigs  
10 were performed to further remove contaminating bacteria. K-mer (k = 21) based methods were  
11 used on filtered reads to estimate genome size of *G. pyriformis* using jellyfish 1.1.12 [62] and  
12 plotted in GenomeScope 2.0 [63] (see figure of “K-mer distribution of filtered illumina genomic  
13 reads of *G. pyriformis*” in Data S1).

#### 14 ***Genome annotation***

15 Protein coding genes were predicted using Funannotate V1.7.4  
16 (<https://funannotate.readthedocs.io/>) [DOI:10.5281/zenodo.3679386], which automates gene  
17 prediction. Assembled transcripts using Trinity [64] and Rnaseq reads mapped bam file were  
18 used as transcript evidence for gene call.

19 Transposable elements were predicted using TransposonPSI [65]. Repeat sequences were first  
20 identified using RepeatModeler [66] with multiple numbers of iterations. The iteration with the  
21 most number of repeats were then used for soft-masking the genome with REPEATMASKER  
22 (open 4.0.646) [16]. Output files generated from above procedures were used to identify repeat  
23 along the assembly. The completeness of genome assembly was assessed with BUSCO version  
24 2.0 [67] with default parameters using the fungal gene dataset [fungi\_odb9] ( see table of  
25 “Genome completeness using Busco” in Data S1).

26 Putative gene functions were identified using Diamond BLASTX [59]. Pfam domain analysis  
27 were performed using Pfamscan [68,69] (Figure S4) (see table “Pfam domain counts from  
28 genomes of mucoromycota used in this study” and “pvalues calculated by fisher’s test” in Data  
29 S1) and Carbohydrate-active enzymes (CAZYme) were identified using the dbCAN CAZY  
30 database [70] (Table S2) (see table of “Presence and absence of plant and fungal cell wall

1 related CAZymes” in Data S1. Putative CAZymes were further verified through comparisons of  
2 data from Morin et al 2019 [17] (Table S1). Secretory proteins were identified using previously  
3 published pipelines [71,72] (see table of “Secretome proteins from genomes” in Data S1), and  
4 effectors were identified using EffectorP 2.0 [73] ( see table of “effector prediction on *G.*  
5 *Pyriiformis* genome” in Data S1). The putative MAT loci of *Paraglomus* sp., and *R. irregularis*  
6 were identified by BLAST search procedures (see figure “Transcriptional directions of the  
7 putative AMF mating-type locus” and table “Identification of Meiosis specific genes” in Data  
8 S1). The tests for a two-evolutionary rates analysis was performed in part by measuring  
9 intergenic distance among genes in genome using a R script [22] ( see figure of “Heatmap  
10 showing evidence for a “one speed genome” in *Geosiphon Pyriiformis*” in Data S1). For this  
11 study, published genomes of additional Glomeromycotina, Mortierellomycotina, and  
12 Mucoromycotina were downloaded from JGI portal MycoCosm database [74,75] [DOI:  
13 10.1093/nar/gkt1183].

#### 14 ***SNP calling and allele frequency analysis***

15 Filtered Mate Pair and Paired End reads were mapped onto the assembled *G. pyriiformis* genome  
16 using the BWA-MEM v 0.7.17 algorithm [76] and sorted into a BAM file using samtools (v 1.9)  
17 [76]. Variants were called using FREEBAYES v1.2.0 [77] and filtered using vcftools [78].  
18 Filtering cutoffs and procedures were as described by in Ropars et al. 2016 and Morin et al. 2019  
19 [17,21] (see figure “Genome wide allele frequency of *G. Pyriiformis* shows it is a homokaryon” .  
20 Quality filtered variants and SNPs which passed filtering were used for constructing allele  
21 frequency plot using a custom R script (code available in GitHub repository).

#### 22 ***Phylogenetic analysis and molecular dating***

23 The phylogenomic analyses employed a set of 434 generally conserved and single-copy proteins  
24 in fungi (data available at DOI: 10.5281/zenodo.1413687), which were developed through efforts  
25 of the 1000 fungal genomes project and provided in the Joint Genome Institute MycoCosm site  
26 [1,75,79]. Profile-Hidden-Markov-Models of these markers were searched in the *Geosiphon*  
27 predicted protein sequences using HMMER3 (v3.1b2) [68] and recovered 393 homologs (out of  
28 the 434) in total. The 434 markers in 45 included fungal genomes were further collapsed into 57  
29 partitions using a greedy search embedded in PartitionFinder v.2.1.1 for consistent phylogenetic  
30 signals [80]. Phylogenetic trees were produced using the PHYling pipeline (data available DOI:

1 10.5281/zenodo.1257002) and with maximum likelihood method implemented in IQ-TREE  
2 (v.1.7-beta9) [81]. Concordance factors across the tree were calculated using the package  
3 implemented in IQ-TREE.

4 The divergence time of *Geosiphon* sp. from the clade of "*Ambispora leptoticha* and *Paraglomus*  
5 *occultum*" was estimated using the R8S v1.81 [82] with the phylogenetic tree reconstructed from  
6 the earlier step. We employed five calibration constraints to calibrate the tree, including the  
7 crown groups of Fungi (1100 MYA) [83], Dikarya (772 MYA) [83], Chytridiomycota (>573  
8 MYA) [84–86], the MRCA of Chytridiomycota and terrestrial fungi (>750 MYA) [84–86], and  
9 Glomeromycotina (>460 MYA) [11]. The divergence time of each clade was inferred using the  
10 Langley-Fitch method with Powell algorithm [87–89].

### 11 **Alternative topology test and dating analyses**

12 To test the likelihood of other possible phylogenetic placements of *G. pyriformis*, we first  
13 reconstructed the associated phylogenetic trees using constraint tree topology as illustrated in  
14 Figure S2 via “-g” option of the IQTREE package (iqtree-1.7-beta9) [81] ( see table “Results of  
15 alternative tree topology test for phylogenetic tree” in Data S1). We then compared our best tree  
16 (shown as Figure 3) with alternative topologies to compute the log-likelihoods of the trees using  
17 Kishino-Hasegawa test, Shimodaira-Hasegawa test, expected likelihood weight, and  
18 approximately unbiased test via “-zb” and “-au” parameters in IQTREE [90–94]. All tests were  
19 performed with 10,000 resampling estimated log-likelihood (RELL) method for reliable results.  
20 The best-fit substitution models for the genome-scale data matrix were estimated using  
21 ModelFinder implemented in IQTREE package [95].

### 22 ***Detection of putative horizontal gene transfers***

23 To identify genes in *Geosiphon* that have potential origin in cyanobacteria, we compared the  
24 *Geosiphon* sp. genome to the available fungal and cyanobacterial genomes. To highlight  
25 potential HGT genes, we used a Python script (available in github repository) [96] to filter out  
26 genome component in *G. pyriformis* with higher similarity score to cyanobacteria than any fungi,  
27 excluding the *G. pyriformis* itself (see figure of “ Phylogenetic tree showing evidence of  
28 horizontal gene transfer in selenium binding protein” and “Phylogenetic tree showing evidence  
29 of horizontal gene transfer in Molybdenum cofactor carrier” in Data S1) (see table of “HGT  
30 containing genes with pfam domains” in Data S1) .



## 1 ***Gene orthology and evolution of symbiotic specific genes***

2 Orthogroups resulting from the OrthoFinder run were parsed using a custom Python script. To be  
3 retained, an orthogroup had to fill the following conditions: any sequence from the non-AMS  
4 fungi, at least one sequence of *Geosiphon pyriformis*, one sequence of either *Rhizophagus*  
5 *irregularis* or *Rhizophagus cerebriforme* and at least one sequence of *Gigaspora rosea* or  
6 *Diversispora epigaea* (see table of “Functions of genes which are gained in orthogroups” in Data  
7 S1). Reciprocally, orthogroups that could correspond to gene losses in the AMS fungi were  
8 extracted by retaining orthogroups with no sequences of AMS fungi and at least one sequence of  
9 each non-AMF fungi (Table S3) and (see table of “Orthogroups lost in the MRCA of  
10 Glomeromycotina” and “Missing MRCA genes across glomeromycotina core genes” in Data  
11 S1).

12 Orthogroups showing evidence of regulation in symbiotic conditions in *R. irregularis* were  
13 subjected to Maximum Likelihood (ML) analysis [92,97] to check for the absence of non-AMS  
14 species. First, proteins contained in orthogroups were searched against the nine proteomes of 9  
15 distinct species using the BLASTp+ v2.9.0 [98] with default parameters and an e-value threshold  
16 fixed at 1e-05 (threshold was set to 1e-03 when no non-AMS species sequences were identified).  
17 Then, proteins were aligned using MUSCLE v3.8.31 [99] with default parameters and resulting  
18 alignment trimmed to remove positions with more than 80% of gaps using trimAl v1.4rev22  
19 [100]. Prior to ML reconstruction, best fitting evolution model was tested using ModelFinder  
20 [95] and then ML analysis was performed using IQ-TREE v1.6.1 [81] with 10,000 replicates of  
21 SH-aLRT. Trees were visualized and annotated with the iTOL platform v5.5 [101,102] (Figure  
22 S3).

23 After first round of phylogeny, orthogroups showing an AMF specific pattern were blasted  
24 against the full MycoCosm database (1565 proteomes, last accessed: 03/01/2020) to confirm the  
25 AMF-specific pattern and a phylogenetic analysis was performed following the procedure  
26 described above.

## 27 ***Differential expression analysis and combination of expression data to orthogroups***

28 Expression data of *Rhizophagus irregularis* in four conditions were used to select orthogroups  
29 containing gene significantly deregulated in symbiosis for further analysis. Paired-end reads  
30 were trimmed and fragments mapped onto Rhiir2\_1 genome assembly of *R. irregularis*

1 ([https://genome.jgi.doe.gov/Rhiir2\\_1/](https://genome.jgi.doe.gov/Rhiir2_1/)). Stringent settings of mapping were used (similarity and  
2 length read mapping criteria at 98% and 95%, respectively). Genes differentially expressed  
3 (DEG) in planta compared to extraradical mycelium were identified after EdgeR [103]  
4 normalization with a false discovery rate (FDR) correction using CLC Genomic Workbench  
5 (Qiagen). We retained genes showing an expression > 2- or < -2-fold times in planta compared to  
6 extraradical hyphae (FDR  $\leq$  0.05). Sets of 2683, 2518 and 2410 DEG were found in *M.*  
7 *truncatula*, *B. distachyon* and *L. cruciata* respectively (see table “Upregulation of ortholog genes  
8 from Rhiir2 genome” and “Downregulation of ortholog genes from Rhiir2 genome” in Data S1).  
9 Detailed information on the data are available at the National Center for Biotechnology  
10 Information (NCBI) Gene Expression Omnibus (GEO) portal (accession no GSE67926). The  
11 analysis performed on RNA-seq data from arbuscocytes in *M. truncatula* [36] presented 6359  
12 DEG.

### 13 **Quantification and Statistical analysis**

14 Statistical analysis and graphs were generated using R studio version 4.0.1 (2020-06-06)

### 15 **References**

- 16 1. Spatafora, J.W., Chang, Y., Benny, G.L., Lazarus, K., Smith, M.E., Berbee, M.L., Bonito,  
17 G., Corradi, N., Grigoriev, I., Gryganskyi, A., *et al.* (2016). A phylum-level phylogenetic  
18 classification of zygomycete fungi based on genome-scale data. *Mycologia* 108, 1028–  
19 1046.
- 20 2. Parniske, M. (2008). Arbuscular mycorrhiza: The mother of plant root endosymbioses.  
21 *Nat. Rev. Microbiol.* 6, 763–775.
- 22 3. Brundrett, M.C. (2009). Mycorrhizal associations and other means of nutrition of vascular  
23 plants: Understanding the global diversity of host plants by resolving conflicting  
24 information and developing reliable means of diagnosis. *Plant Soil* 320, 37–77.
- 25 4. Smith, S., and Read, D. (2008). *Mycorrhizal Symbiosis*.
- 26 5. Remy, W., Taylor, T.N., Hass, H., and Kerp, H. (1994). Four hundred-million-year-old  
27 vesicular arbuscular mycorrhizae. *Proc. Natl. Acad. Sci. U. S. A.* 91, 11841–11843.
- 28 6. Strullu-Derrien, C., Selosse, M.A., Kenrick, P., and Martin, F.M. (2018). The origin and  
29 evolution of mycorrhizal symbioses: from palaeomycology to phylogenomics. *New*

- 1       Phytol. 220, 1012–1030.
- 2   7.    Delaux, P.M., Radhakrishnan, G. V., Jayaraman, D., Cheema, J., Malbreil, M., Volkening,  
3       J.D., Sekimoto, H., Nishiyama, T., Melkonian, M., Pokorny, L., *et al.* (2015). Algal  
4       ancestor of land plants was preadapted for symbiosis. *Proc. Natl. Acad. Sci. U. S. A.* 112,  
5       13390–13395.
- 6   8.    Radhakrishnan, G. V., Keller, J., Rich, M.K., Vernié, T., Mbadanga Mbadanga, D.L.,  
7       Vigneron, N., Cottret, L., Clemente, H.S., Libourel, C., Cheema, J., *et al.* (2020). An  
8       ancestral signalling pathway is conserved in intracellular symbioses-forming plant  
9       lineages. *Nat. Plants* 6, 280–289.
- 10  9.    Beaudet, D., Chen, E.C.H., Mathieu, S., Yildirim, G., Ndikumana, S., Dalpé, Y., Séguin,  
11       S., Farinelli, L., Stajich, J.E., and Corradi, N. (2018). Ultra-low input transcriptomics  
12       reveal the spore functional content and phylogenetic affiliations of poorly studied  
13       arbuscular mycorrhizal fungi. *DNA Res.* 25, 217–227.
- 14  10.   Gehrig, H., Schüßler, A., and Kluge, M. (1996). Geosiphon pyriforme, a fungus forming  
15       endocytobiosis with Nostoc (cyanobacteria), is an ancestral member of the glomales:  
16       Evidence by SSU rRNA analysis. *J. Mol. Evol.* 43, 71–81.
- 17  11.   Redecker, D. (2000). Glomalean Fungi from the Ordovician. *Science* 289, 1920–1921.
- 18  12.   Schüßler, A., and Walker, C. (2011). 7 Evolution of the ‘Plant-Symbiotic’ Fungal Phylum,  
19       Glomeromycota. In *Evolution of Fungi and Fungal-Like Organisms*, pp. 163–185.
- 20  13.   Schüßler, A., Schnepf, E., Mollenhauer, D., and Kluge, M. (1995). The fungal bladders of  
21       the endocyanosis *Geosiphon pyriforme*, a *Glomus*-related fungus: cell wall permeability  
22       indicates a limiting pore radius of only 0.5 nm. *Protoplasma* 185, 131–139.
- 23  14.   Sun, X., Chen, W., Ivanov, S., MacLean, A.M., Wight, H., Ramaraj, T., Mudge, J.,  
24       Harrison, M.J., and Fei, Z. (2019). Genome and evolution of the arbuscular mycorrhizal  
25       fungus *Diversispora epigaea* (formerly *Glomus versiforme*) and its bacterial  
26       endosymbionts. *New Phytol.* 221, 1556–1573.
- 27  15.   Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P.,  
28       Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner.  
29       *Bioinformatics* 29, 15–21.

- 1 16. Smit, A.F.A., Hubley, R., and Green, P. (2010). RepeatMasker Open-3.0. 1996-2010. Inst.  
2 Syst. Biol.
- 3 17. Morin, E., Miyauchi, S., San Clemente, H., Chen, E.C.H., Pelin, A., de la Providencia, I.,  
4 Ndikumana, S., Beaudet, D., Hainaut, M., Drula, E., *et al.* (2019). Comparative genomics  
5 of *Rhizophagus irregularis*, *R. cerebriforme*, *R. diaphanus* and *Gigaspora rosea* highlights  
6 specific genetic features in Glomeromycotina. *New Phytol.* 222, 1584–1598.
- 7 18. Tisserant, E., Malbreil, M., Kuo, A., Kohler, A., Symeonidi, A., Balestrini, R., Charron,  
8 P., Duensing, N., Frei Dit Frey, N., Gianinazzi-Pearson, V., *et al.* (2013). Genome of an  
9 arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proc. Natl.*  
10 *Acad. Sci. U. S. A.* 110, 20117–20122.
- 11 19. Mathieu, S., Cusant, L., Roux, C., and Corradi, N. (2018). Arbuscular mycorrhizal fungi:  
12 intraspecific diversity and pangenomes. *New Phytol.* 220, 1129–1134.
- 13 20. Chen, E.C.H., Morin, E., Beaudet, D., Noel, J., Yildirim, G., Ndikumana, S., Charron, P.,  
14 St-Onge, C., Giorgi, J., Krüger, M., *et al.* (2018). High intraspecific genome diversity in  
15 the model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*. *New Phytol.* 220,  
16 1161–1171. Available at: <http://doi.wiley.com/10.1111/nph.14989> [Accessed September  
17 9, 2019].
- 18 21. Ropars, J., Toro, K.S., Noel, J., Pelin, A., Charron, P., Farinelli, L., Marton, T., Krüger,  
19 M., Fuchs, J., Brachmann, A., *et al.* (2016). Evidence for the sexual origin of  
20 heterokaryosis in arbuscular mycorrhizal fungi. *Nat. Microbiol.* 1.
- 21 22. Dong, S., Raffaele, S., and Kamoun, S. (2015). The two-speed genomes of filamentous  
22 pathogens: Waltz with plants. *Curr. Opin. Genet. Dev.* 35, 57–65.
- 23 23. Mathu Malar, C., Yuzon, J.D., Das, S., Das, A., Panda, A., Ghosh, S., Tyler, B.M.,  
24 Kasuga, T., and Tripathy, S. (2019). Haplotype-phased genome assembly of virulent  
25 phytophthora ramorum isolate ND886 facilitated by long-read sequencing reveals effector  
26 polymorphisms and copy number variation. *Mol. Plant-Microbe Interact.* 32, 1047–1060.
- 27 24. Vandenkoornhuys, P., Leyval, C., and Bonnin, I. (2001). High genetic diversity in  
28 arbuscular mycorrhizal fungi: Evidence for recombination events. *Heredity (Edinb).* 87,  
29 243–253.

- 1 25. Chen, E.C.C.H., Mathieu, S., Hoffrichter, A., Sedziewska-Toro, K., Peart, M., Pelin, A.,  
2 Ndikumana, S., Ropars, J., Dreissig, S., Fuchs, J., *et al.* (2019). Single nucleus sequencing  
3 reveals evidence of inter-nucleus recombination in arbuscular mycorrhizal fungi. *Elife* 7,  
4 1–17.
- 5 26. Croll, D., and Sanders, I.R. (2009). Recombination in *Glomus intraradices*, a supposed  
6 ancient asexual arbuscular mycorrhizal fungus. *BMC Evol. Biol.* 9.
- 7 27. Krüger, M., Krüger, C., Walker, C., Stockinger, H., and Schüßler, A. (2012). Phylogenetic  
8 reference data for systematics and phylotaxonomy of arbuscular mycorrhizal fungi from  
9 phylum to species level. *New Phytol.* 193, 970–984.
- 10 28. Emms, D.M., and Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for  
11 comparative genomics. *Genome Biol.* 20.
- 12 29. Corradi, N., and Lildhar, L. (2012). Meiotic genes in the arbuscular mycorrhizal fungi.  
13 *Commun. Integr. Biol.* 5, 187–189.
- 14 30. Halary, S., Malik, S.B., Lildhar, L., Slamovits, C.H., Hijri, M., and Corradi, N. (2011).  
15 Conserved meiotic machinery in *Glomus* spp., a putatively ancient asexual fungal lineage.  
16 *Genome Biol. Evol.* 3, 950–958.
- 17 31. Idnurm, A., Walton, F.J., Floyd, A., and Heitman, J. (2008). Identification of the sex  
18 genes in an early diverged fungus. *Nature* 451, 193–196.
- 19 32. Wong, S., Fares, M.A., Zimmermann, W., Butler, G., and Wolfe, K.H. (2003). Evidence  
20 from comparative genomics for a complete sexual cycle in the “asexual” pathogenic yeast  
21 *Candida glabrata*. *Genome Biol.* 4.
- 22 33. Gaude, N., Bortfeld, S., Duensing, N., Lohse, M., and Krajinski, F. (2012). Arbuscule-  
23 containing and non-colonized cortical cells of mycorrhizal roots undergo extensive and  
24 specific reprogramming during arbuscular mycorrhizal development. *Plant J.* 69, 510–528.
- 25 34. An, J., Zeng, T., Ji, C., de Graaf, S., Zheng, Z., Xiao, T.T., Deng, X., Xiao, S., Bisseling,  
26 T., Limpens, E., *et al.* (2019). A *Medicago truncatula* SWEET transporter implicated in  
27 arbuscule maintenance during arbuscular mycorrhizal symbiosis. *New Phytol.*
- 28 35. Zeng, T., Holmer, R., Hontelez, J., te Lintel-Hekkert, B., Marufu, L., de Zeeuw, T., Wu,

- 1 F., Schijlen, E., Bisseling, T., and Limpens, E. (2018). Host- and stage-dependent  
2 secretome of the arbuscular mycorrhizal fungus *Rhizophagus irregularis*. *Plant J.* *94*, 411–  
3 425.
- 4 36. Fiorilli, V., Catoni, M., Miozzi, L., Novero, M., Accotto, G.P., and Lanfranco, L. (2009).  
5 Global and cell-type gene expression profiles in tomato plants colonized by an arbuscular  
6 mycorrhizal fungus. *New Phytol.*
- 7 37. Sheng, M., Chen, X., Zhang, X., Hamel, C., Cui, X., Chen, J., Chen, H., and Tang, M.  
8 (2017). Changes in arbuscular mycorrhizal fungal attributes along a chronosequence of  
9 black locust (*Robinia pseudoacacia*) plantations can be attributed to the plantation-induced  
10 variation in soil properties. *Sci. Total Environ.*
- 11 38. Berruti, A., Demasi, S., Lumini, E., Kobayashi, N., Scariot, V., and Bianciotto, V. (2017).  
12 Wild *Camellia japonica* specimens in the Shimane prefecture (Japan) host previously  
13 undescribed AMF diversity. *Appl. Soil Ecol.*
- 14 39. Krüger, C., Kohout, P., Janoušková, M., Püschel, D., Frouz, J., and Rydlová, J. (2017).  
15 Plant communities rather than soil properties structure arbuscular mycorrhizal fungal  
16 communities along primary succession on a mine spoil. *Front. Microbiol.*
- 17 40. Temina, M., Rezankova, H., Rezanka, T., and Dembitsky, V.M. (2007). Diversity of the  
18 fatty acids of the *Nostoc* species and their statistical analysis. *Microbiol. Res.* *162*, 308–  
19 321.
- 20 41. Patel, V.K., Sundaram, S., Patel, A.K., and Kalra, A. (2018). Characterization of Seven  
21 Species of Cyanobacteria for High-Quality Biomass Production. *Arab. J. Sci. Eng.*
- 22 42. López-Rosales, A.R., Ancona-Canché, K., Chavarria-Hernandez, J.C., Barahona-Pérez,  
23 F., Toledano-Thompson, T., Garduño-Solórzano, G., López-Adrian, S., Canto-Canché, B.,  
24 Polanco-Lugo, E., and Valdez-Ojeda, R. (2019). Fatty acids, hydrocarbons and terpenes of  
25 nannochloropsis and nannochloris isolates with potential for biofuel production. *Energies.*
- 26 43. Steinhoff, F.S., Karlberg, M., Graeve, M., and Wulff, A. (2014). Cyanobacteria in  
27 Scandinavian coastal waters - A potential source for biofuels and fatty acids? *Algal Res.*
- 28 44. Rich, M.K., Nouri, E., Courty, P.E., and Reinhardt, D. (2017). Diet of Arbuscular  
29 Mycorrhizal Fungi: Bread and Butter? *Trends Plant Sci.* *22*, 652–660.

- 1 45. Luginbuehl, L.H., Menard, G.N., Kurup, S., Van Erp, H., Radhakrishnan, G. V.,  
2 Breakspear, A., Oldroyd, G.E.D., and Eastmond, P.J. (2017). Fatty acids in arbuscular  
3 mycorrhizal fungi are synthesized by the host plant. *Science* (80-. ). 356, 1175–1178.
- 4 46. Cordeiro, R.S., Vaz, I.C.D., Magalhães, S.M.S., and Barbosa, F.A.R. (2017). Effects of  
5 nutritional conditions on lipid production by cyanobacteria. *An. Acad. Bras. Cienc.*
- 6 47. Keymer, A., Pimprikar, P., Wewer, V., Huber, C., Brands, M., Bucerius, S.L., Delaux,  
7 P.M., Klingl, V., von Röpenack-Lahaye, E., Wang, T.L., *et al.* (2017). Lipid transfer from  
8 plants to arbuscular mycorrhiza fungi. *Elife* 6.
- 9 48. Torres-Cortés, G., Ghignone, S., Bonfante, P., and Schüßler, A. (2015). Mosaic genome of  
10 endobacteria in arbuscular mycorrhizal fungi: Transkingdom gene transfer in an ancient  
11 mycoplasma-fungus association. *Proc. Natl. Acad. Sci. U. S. A.* 112, 7785–7790.
- 12 49. Rensing, S.A. (2018). Great moments in evolution: the conquest of land by plants. *Curr.*  
13 *Opin. Plant Biol.* 42, 49–54.
- 14 50. Morris, J.L., Puttick, M.N., Clark, J.W., Edwards, D., Kenrick, P., Pressel, S., Wellman,  
15 C.H., Yang, Z., Schneider, H., and Donoghue, P.C.J. (2018). The timescale of early land  
16 plant evolution. *Proc. Natl. Acad. Sci. U. S. A.* 115, E2274–E2283.
- 17 51. Genre, A., Chabaud, M., Balzergue, C., Puech-Pagès, V., Novero, M., Rey, T., Fournier,  
18 J., Rochange, S., Bécard, G., Bonfante, P., *et al.* (2013). Short-chain chitin oligomers from  
19 arbuscular mycorrhizal fungi trigger nuclear Ca<sup>2+</sup> spiking in *Medicago truncatula* roots  
20 and their production is enhanced by strigolactone. *New Phytol.* 198, 190–202.
- 21 52. Maillet, F., Poinso, V., André, O., Puech-Pagés, V., Haouy, A., Gueunier, M., Cromer,  
22 L., Giraudet, D., Formey, D., Niebel, A., *et al.* (2011). Fungal lipochitooligosaccharide  
23 symbiotic signals in arbuscular mycorrhiza. *Nature* 469, 58–64.
- 24 53. Schüßler, A., and Wolf, E. (2005). *Geosiphon pyriformis*—a Glomeromycotan Soil  
25 Fungus Forming Endosymbiosis with Cyanobacteria. In *In Vitro Culture of Mycorrhizas*,  
26 pp. 271–289.
- 27 54. Mollenhauer, D., and Mollenhauer, R. (1988). *Geosiphon* cultures ahead. *Endocytobios*  
28 *Cell Res* 5, 69–73.

- 1 55. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for  
2 Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- 3 56. Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017). MetaSPAdes: A  
4 new versatile metagenomic assembler. *Genome Res.* 27, 824–834.
- 5 57. Alneberg, J., Bjarnason, B.S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L.,  
6 Loman, N.J., Andersson, A.F., and Quince, C. (2014). Binning metagenomic contigs by  
7 coverage and composition. *Nat. Methods* 11, 1144–1146.
- 8 58. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local  
9 alignment search tool. *J Mol Biol* 215, 403–410.
- 10 59. Buchfink, B., Xie, C., and Huson, D.H. (2014). Fast and sensitive protein alignment using  
11 DIAMOND. *Nat. Methods* 12, 59–60.
- 12 60. Kent, W.J. (2002). BLAT---The BLAST-Like Alignment Tool.
- 13 61. Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S.L., and Yorke, J.A. (2013).  
14 The MaSuRCA genome assembler. *Bioinformatics* 29, 2669–2677.
- 15 62. Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel  
16 counting of occurrences of k-mers. *Bioinformatics* 27, 764–770.
- 17 63. Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J.,  
18 and Schatz, M.C. (2017). GenomeScope: Fast reference-free genome profiling from short  
19 reads. In *Bioinformatics*, pp. 2202–2204.
- 20 64. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J.,  
21 Couger, M.B., Eccles, D., Li, B., Lieber, M., *et al.* (2013). De novo transcript sequence  
22 reconstruction from RNA-seq using the Trinity platform for reference generation and  
23 analysis. *Nat. Protoc.* 8, 1494–1512.
- 24 65. Hass, B. (2010). TransposonPSI: An application of PSI-Blast to mine (retro-)transposon  
25 ORF homologies. Broad Institute, Cambridge, MA, USA.
- 26 66. Smit, A., and Hubley, R. (2013). RepeatModeler. RepeatModeler.
- 27 67. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E.M.  
28 (2015). BUSCO: Assessing genome assembly and annotation completeness with single-



- 1 copy orthologs. *Bioinformatics* 31, 3210–3212.
- 2 68. Hancock, J.M., Zvelebil, M.J., Hancock, J.M., and Bishop, M.J. (2004). HMMer. In  
3 Dictionary of Bioinformatics and Computational Biology.
- 4 69. Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A.,  
5 Hetherington, K., Holm, L., Mistry, J., *et al.* (2014). Pfam: The protein families database.  
6 *Nucleic Acids Res.*
- 7 70. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014).  
8 The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42.
- 9 71. Kamel, L., Tang, N., Malbreil, M., San Clemente, H., Le Marquer, M., Roux, C., and dit  
10 Frey, N.F. (2017). The comparison of expressed candidate secreted proteins from two  
11 arbuscular mycorrhizal fungi unravels common and specific molecular tools to invade  
12 different host plants. *Front. Plant Sci.* 8.
- 13 72. Pellegrin, C., Morin, E., Martin, F.M., and Veneault-Fourrey, C. (2015). Comparative  
14 analysis of secretomes from ectomycorrhizal fungi with an emphasis on small-secreted  
15 proteins. *Front. Microbiol.* 6.
- 16 73. Sperschneider, J., Dodds, P.N., Gardiner, D.M., Singh, K.B., and Taylor, J.M. (2018).  
17 Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Mol.*  
18 *Plant Pathol.* 19, 2094–2110.
- 19 74. Grigoriev, I. V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., Kuo, A.,  
20 Minovitsky, S., Nikitin, R., Ohm, R.A., *et al.* (2012). The Genome Portal of the  
21 Department of Energy Joint Genome Institute. *Nucleic Acids Res.* 40.
- 22 75. Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otilar, R., Riley, R., Salamov,  
23 A., Zhao, X., Korzeniewski, F., *et al.* (2014). MycoCosm portal: Gearing up for 1000  
24 fungal genomes. *Nucleic Acids Res.* 42.
- 25 76. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-  
26 Wheeler transform. *Bioinformatics* 25, 1754–1760.
- 27 77. Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read  
28 sequencing. *arXiv Prepr. arXiv1207.3907*. Available at: <http://arxiv.org/abs/1207.3907>.

- 1 78. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A.,  
2 Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., *et al.* (2011). The variant call  
3 format and VCFtools. *Bioinformatics* 27, 2156–2158.
- 4 79. Stajich, J.E. (2017). Fungal Genomes and Insights into the Evolution of the Kingdom. *The*  
5 *Fungal Kingdom*, 619–633.
- 6 80. Lanfear, R., Calcott, B., Ho, S.Y.W., and Guindon, S. (2012). PartitionFinder: Combined  
7 selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol.*  
8 *Biol. Evol.* 29, 1695–1701.
- 9 81. Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A fast  
10 and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol.*  
11 *Biol. Evol.* 32, 268–274.
- 12 82. Sanderson, M.J. (2003). r8s: Inferring absolute rates of molecular evolution and  
13 divergence times in the absence of a molecular clock. *Bioinformatics* 19, 301–302.
- 14 83. Parfrey, L.W., Lahr, D.J.G., Knoll, A.H., and Katz, L.A. (2011). Estimating the timing of  
15 early eukaryotic diversification with multigene molecular clocks. *Proc. Natl. Acad. Sci. U.*  
16 *S. A.* 108, 13624–13629.
- 17 84. Lutzoni, F., Nowak, M.D., Alfaro, M.E., Reeb, V., Miadlikowska, J., Krug, M., Arnold,  
18 A.E., Lewis, L.A., Swofford, D., Hibbett, D., *et al.* (2018). Contemporaneous radiations of  
19 fungi and plants linked to symbiosis. *Nat. Commun.* 9, 1–11.
- 20 85. Wang, Y., White, M.M., and Moncalvo, J.-M.M. (2019). Diversification of the gut fungi  
21 *Smittium* and allies (Harpellales) co-occurred with the origin of complete metamorphosis  
22 of their symbiotic insect hosts (lower Diptera). *Mol. Phylogenet. Evol.* 139, 106550.
- 23 86. Chang, Y., Wang, S., Sekimoto, S., Aerts, A., Choi, C., Clum, A., LaButti, K., Lindquist,  
24 E., Ngan, C.Y., Ohm, R.A., *et al.* (2015). Phylogenomic analyses indicate that early fungi  
25 evolved digesting cell walls of algal ancestors of land plants. *Genome Biol. Evol.* 7, 1590–  
26 1601.
- 27 87. Langley, C.H., and Fitch, W.M. (1974). An examination of the constancy of the rate of  
28 molecular evolution. *J. Mol. Evol.* 3, 161–177.

- 1 88. Gill, P.E., Murray, W., and Wright, M.H. (1981). Practical optimization (New York:  
2 Academic Press).
- 3 89. Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. (1992). Numerical  
4 recipes in C 2nd ed. (New York: Cambridge University Press).
- 5 90. Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection.  
6 *Syst. Biol.* *51*, 492–508.
- 7 91. Shimodaira, H., and Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with  
8 applications to phylogenetic inference. *Mol. Biol. Evol.* *16*, 1114–1116.
- 9 92. Kishino, H., Miyata, T., and Hasegawa, M. (1990). Maximum likelihood inference of  
10 protein phylogeny and the origin of chloroplasts. *J. Mol. Evol.* *31*, 151–160.
- 11 93. Kishino, H., and Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of  
12 the evolutionary tree topologies from DNA sequence .... *J. Mol. Evol.* *29*, 170–179.
- 13 94. Strimmer, K., and Rambaut, A. (2002). Inferring confidence sets of possibly misspecified  
14 gene trees. *Proc. R. Soc. B Biol. Sci.* *269*, 137–142.
- 15 95. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A., and Jermini, L.S.  
16 (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat.*  
17 *Methods* *14*, 587–589.
- 18 96. Wang, Y., White, M.M., Kvist, S., and Moncalvo, J.M. (2016). Genome-Wide Survey of  
19 Gut Fungi (Harpellales) Reveals the First Horizontally Transferred Ubiquitin Gene from a  
20 Mosquito Host. *Mol. Biol. Evol.* *33*, 2544–2554.
- 21 97. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O.  
22 (2010). New algorithms and methods to estimate maximum-likelihood phylogenies:  
23 Assessing the performance of PhyML 3.0. *Syst. Biol.* *59*, 307–321.
- 24 98. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and  
25 Madden, T.L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics* *10*.
- 26 99. Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high  
27 throughput. *Nucleic Acids Res.* *32*, 1792–1797.
- 28 100. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: A tool for

1 automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25,  
2 1972–1973.

3 101. Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): An online tool for  
4 phylogenetic tree display and annotation. *Bioinformatics* 23, 127–128.

5 102. Letunic, I., and Bork, P. (2019). Interactive Tree of Life (iTOL) v4: Recent updates and  
6 new developments. *Nucleic Acids Res.* 47.

7 103. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2009). edgeR: A Bioconductor  
8 package for differential expression analysis of digital gene expression data.  
9 *Bioinformatics* 26, 139–140.

10  
11  
12  
13

14 **Table 1: Summary statistics for genome assembly of sequenced *G. pyriformis* and other**  
15 **species from Glomeromycotina and selected Mucoromycota used in this study**

Genomes	Assembly size	No of scaffolds	Scaffold N50	Largest scaffold (Kb)	Total Gap%	Repeat %	Busco completeness %	GC %
<i>Geosiphon Pyriformis</i>	129	795	703	2733.91	0.023	64.35	96.2	29.25
<i>Gigaspora rosea</i> VI.0	597.95	7526	734	1204.75	7.92	63.44	97.9	28.81
<i>Rhizophagus ceribriforme</i> DAOM227022 VI.0	136.89	2592	266	709.02	17.60	24.77	98.3	26.55

<i>Rhizophagus irregularis</i> DAOM 197198V2.0	136.80	1123	129	1375.86	5.06	26.38	98	27.53
<i>Diversispora versiformis</i> strain IT104	147	731	434	2010.39	0.061	43.6	98.2	25.1
<i>Rhizopus microsporus</i> ATCC11559 VI	25.97	131	8	2782.17	2.41	4.68	98.6	37.48
<i>Mucor Circinelloides</i> CBS 277.49 V2.0	36.59	26	4	6050.25	0.00	20.38	97.2	42.17
<i>Phycomyces Blakesleeanus</i> NRRL1555 V2.0	53.94	80	11	4452.46	1.06	9.74	96.9	35.78
<i>Mortierella Elongata</i> AG-77	49.86	473	31	1526.29	0.30	4.63	99.7	48.05

1

2

3