UC Riverside

UC Riverside Previously Published Works

Title

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

Permalink

https://escholarship.org/uc/item/0xz2n8zv

Journal

Current Biology, 31(7)

ISSN

0960-9822

Authors

Malar C, Mathu
Krüger, Manuela
Krüger, Claudia
et al.

Publication Date

2021-04-01

DOI

10.1016/j.cub.2021.01.058

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at https://creativecommons.org/licenses/by-nc-nd/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

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

Mathu Malar C [1*], Manuela Krüger [2,3,*], Claudia Krüger [2,3*], Yan Wang [4,5], Jason E. Stajich [6], Jean Keller [7], Eric C.H. Chen[1], Gokalp Yildirir [1], Matthew Villeneuve-Laroche[1], Christophe Roux [7], Pierre-Marc Delaux [7] and Nicolas Corradi [1]

[1] Department of Biology, University of Ottawa. Ottawa, Canada.

[2] Institute of Botany, The Czech Academy of Sciences. Průhonice, Czech Republic

[3] Institute of Experimental Botany, The Czech Academy of Science, Prague, Czech Republic

[4] Department of Biological Sciences, University of Toronto Scarborough, Toronto, Canada

[5] Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada

[6] Department of Microbiology & Plant Pathology and Institute for Integrative Genome Biology, University of California–Riverside, Riverside, California. USA

[7] Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, UPS, CNRS 24 Chemin de Borde Rouge-Auzeville, Castanet-Tolosan, France

*contributed equally

Correspondence should be addressed to: ncorradi@uottawa.ca

Abstract

Arbuscular mycorrhizal fungi (AMF, sub-phylum Glomeromycotina) are prominent root symbionts of land plants. This symbiotic association, named arbuscular mycorrhizal symbiosis (AMS), allows the fungus to receive photosynthetically fixed carbohydrates from the plant in exchange of improved nutrients and water uptake by the roots. AMS is considered to have played a key role in the colonization of the land by plants, but how this symbiotic ability emerged is not entirely understood due to lack of genomes from AMF representatives of basal phylogenetic nodes. To address this, we sequenced the genome of the AMF *Geosiphon pyriformis*, a species involved in the only known fungal endosymbiosis with a cyanobacterium. This species is unknown to undergo AMS, yet we find that its genome carries all hallmarks of AMF obligate plant biotrophy, including a reduced set of plant cell wall degrading enzymes and an absence of

genes for the production of sugars and fatty acids. Using comparative genomics and available RNA-seq data from a model species, we identify a set of genes lost in the MRCA of Glomeromycotina, and a putative core gene set differentially regulated during AMS. Overall, our findings indicate that the mechanisms involved in AMS appeared prior to the emergence of Glomeromycotina. These also provide a basis for future research on symbiotic mechanisms in prominent plant mutualists and primary insights into a fungus-cyanobacteria endosymbiosis.

Introduction

Within the continuum of interactions between organisms, from parasitism to mutualism, the symbiosis formed by plants and arbuscular mycorrhizal fungi (AMF) (subphylum Glomeromycotina) [1] is amongst the most prominent [2]. Over 70% of known land plant species have the potential to form this symbiosis (AMS) [3]. AMS plays a major role in extant terrestrial ecosystems as it allows plants to efficiently acquire poorly soluble soil nutrients [4]. During AMS, the fungal partner colonizes the plant roots (or the thalli of root-less plants), and forms small tree-like structures called arbuscules within the plant cell that promote bidirectional nutrient transfers between partners. Thereby, AMF receives photosynthetically fixed carbohydrates from the plant and, in return, the fungus delivers phosphate, nitrogen and other trace elements as well as water to its host. This plant-fungus symbiosis dates back more than 400 million years [5], and is considered as one of the key innovations that allowed the colonization of lands by plants [6] .

Genomic and genetic analyses conducted in diverse plant species have started to shed light on the molecular mechanisms that allowed the evolution of this symbiosis on the host side [7,8]. By contrast, how and when AMS abilities emerged in AMF remains elusive. Comparative phylogenomics could be used to understand the evolution AMS [7,8], however, such approaches require the availability of genome data covering the diversity of AMF lineages. Based on ribosomal RNA (rRNA) phylogenies, AMF are separated into four orders, namely Diversisporales, Glomerales, Archaeosporales and Paraglomerales. To date, only genomes from Diversisporales and Glomerales have been sequenced, while only fragmentary protein coding datasets are available for basal phylogenetic nodes that include the Archaeosporales (*Ambispora leptoticha*) and Paraglomerales (*Paraglomus occultum*) [9].

Among the Archaeosporales, *Geosiphon pyriformis* stands out. This species is the only fungus known to produce endosymbiosis with nitrogen-fixing cyanobacteria (*Nostoc punctiforme*), and it was proposed that this association represent the AMF ancestral state [10–12]. As opposed to other AMF, which penetrate the plant cells to form arbuscules, *G. pyriformis* forms long fungal cells (‘bladders’) that enclose cyanobacteria. Once in the bladder, the cyanobacterium is photosynthetically active and fixes nitrogen, and in exchange the fungus delivers inorganic nutrients and provides water. The bladder membrane, which encloses the cyanobacteria cells, is

an extension of the fungal plasma membrane [14]. The first AMF monosaccharide transporter (GpMST1) was found using the *G. pyriformis-N. punctiforme* endosymbiosis system [15,16], and despite almost four decades of research on this species *G. pyriformis* was never found to form intracellular symbiotic structures such as arbuscules, [15,17,18].

Arguably, *G. pyriformis* represents an ideal candidate to investigate the origin of AMS and the emergence of a unique endosymbiosis. Here, we aimed to advance knowledge in these questions by sequencing the genome of *G. pyriformis* using a re-discovered isolate.

Results

General genome characteristics of Geosiphon pyriformis

The only known culture of *G. pyriformis* was lost over a decade ago (A. Schüßler, pers. comm). In an attempt to rediscover *G. pyriformis*, we searched and identified symbiotic bladders of the *G. pyriformis-N. punctiforme* symbiosis (Figure 1) at the only known stable habitat of this species in the Spessart Mountains near the village of Bieber in Germany [19].

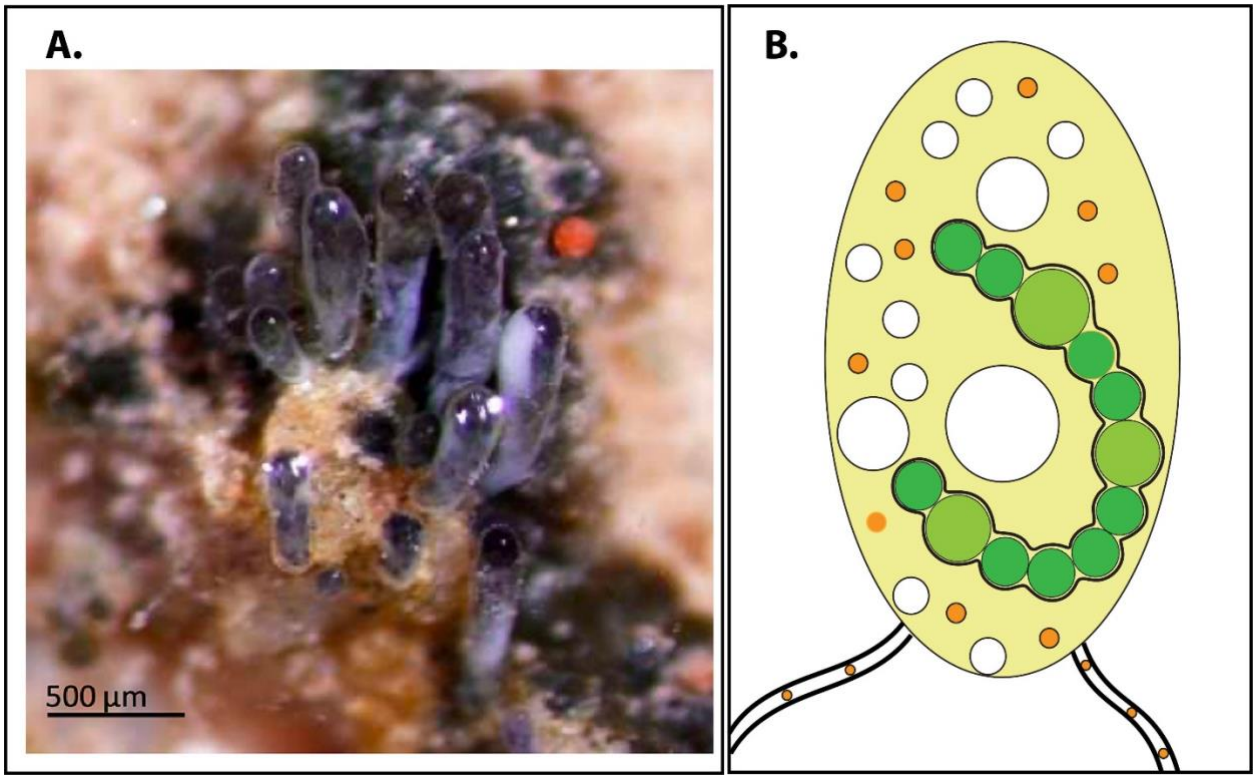


Figure 1: a. Image of *G. pyriformis* bladders in soil from its natural habitat in the Spessart mountains. **b**. Schematic representation of *G. pyriformis* bladders containing *Nostoc* cells (based on M. Kluge 2002). *Nostoc* cells are shown in dark green, and heterocysts (differentiated cell that carries out nitrogen fixation) are shown in light green. Bladders contain *G. pyriformis* nuclei (orange), and several vacuoles (white). Aseptate hyphae reach out and extend from the bladders.

Upon cultivation, we extracted DNA and RNA from active bladders. Total DNA was subjected to 5Kb-mate-pair and 125 bp paired-end Illumina sequencing, producing respectively 47 and 81 million 125 bp paired ends and 5kb mate-pairs reads. *G. pyriformis* reads were identified using a read binning approach recently implemented to assemble the genome of *Diversispora epigea epigaea*[16], and upon identification these were assembled into a 129 MB assembly and 795 scaffolds with an average read coverage of 118X. In parallel, total RNA was subjected to 150 bp paired-end Illumina sequencing. The resulting RNA-seq reads were mapped onto the *G. pyriformis* genome assembly using STAR [20] and used for genome annotation after implementing RepeatMasker[18]. This procedure identified of 24195 genes in *G. pyriformis,* resulting in a BUSCO gene repertoire completeness of 96.2% (3.1% complete duplicated). The gene counts, estimated genome size, and genome statistics are all similar to those of model AMF species [13,21] and are indicative of high genome completeness (Table 1, Supplementary Table 1 and Supplementary Figure 1). SignalP showed that, among all the genes identified in *G. pyriformis,* 365 represent putative secreted proteins (Supplementary Table 2) and 27% of these are candidate effectors (Supplementary Table 3) [10,11]. We also identified 19 putative secreted CAZymes in *G. pyriformis*; in line with numbers found in other AMF species (Supplementary Table 4 and Supplementary Table 5).

AMF genomes carry a substantial fraction of transposable elements (TE) [10,11,25,26], and we found that *G. pyriformis* has undergone similar TE expansions. The expansion of Gypsy transposable elements in *G. pyriformis* is evident in comparison to all other AMF genomes (Figure 2). With regards to TE, we find no evidence that *G. pyriformis* carries a two-speed genome (Supplementary Figure 2) [29]. Two-speed genomes are characterized by the presence of TE-poor and gene dense regions that a clearly separate from others that contain rapidly evolving genomic regions that usually carry less genes, abundant TE, and other repeat elements [30,31].

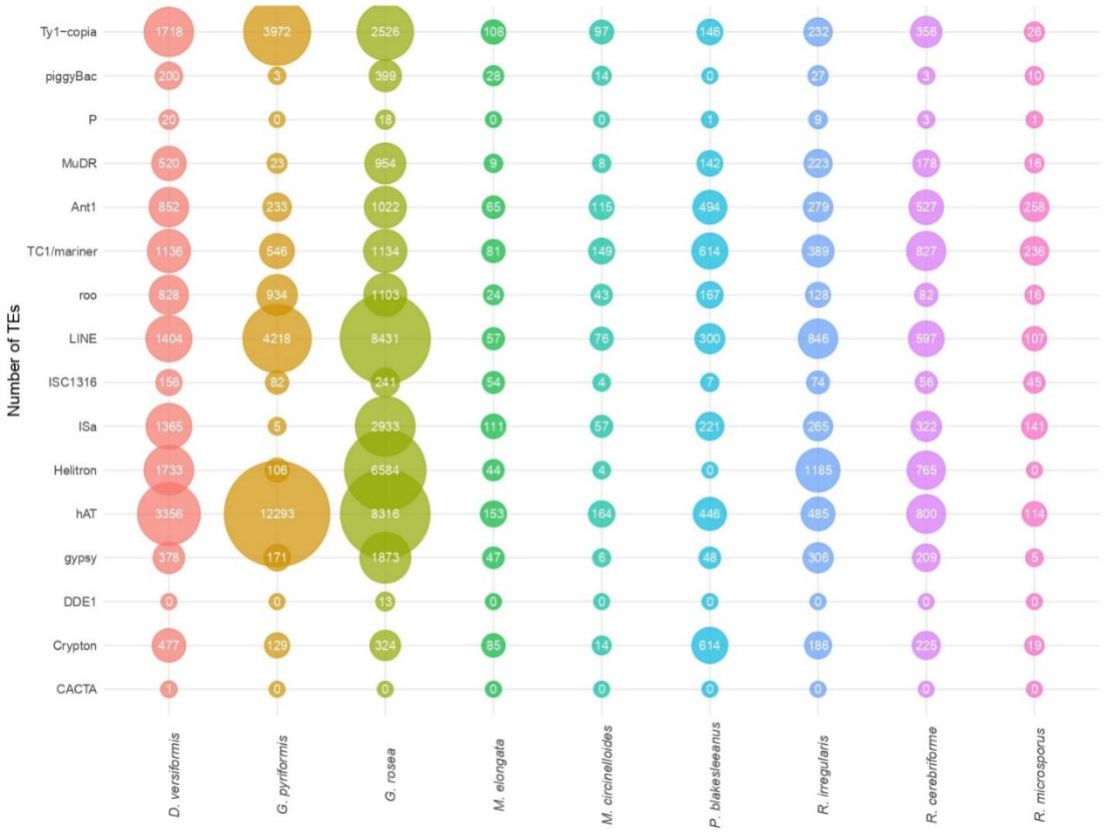


Figure 2 Bubble plot containing all transposable elements in the genomes of glomeromycotina and mucoromycotina genomes used in this study. The figure shows the expansion of Gypsy elements in *G. pyriformis.*

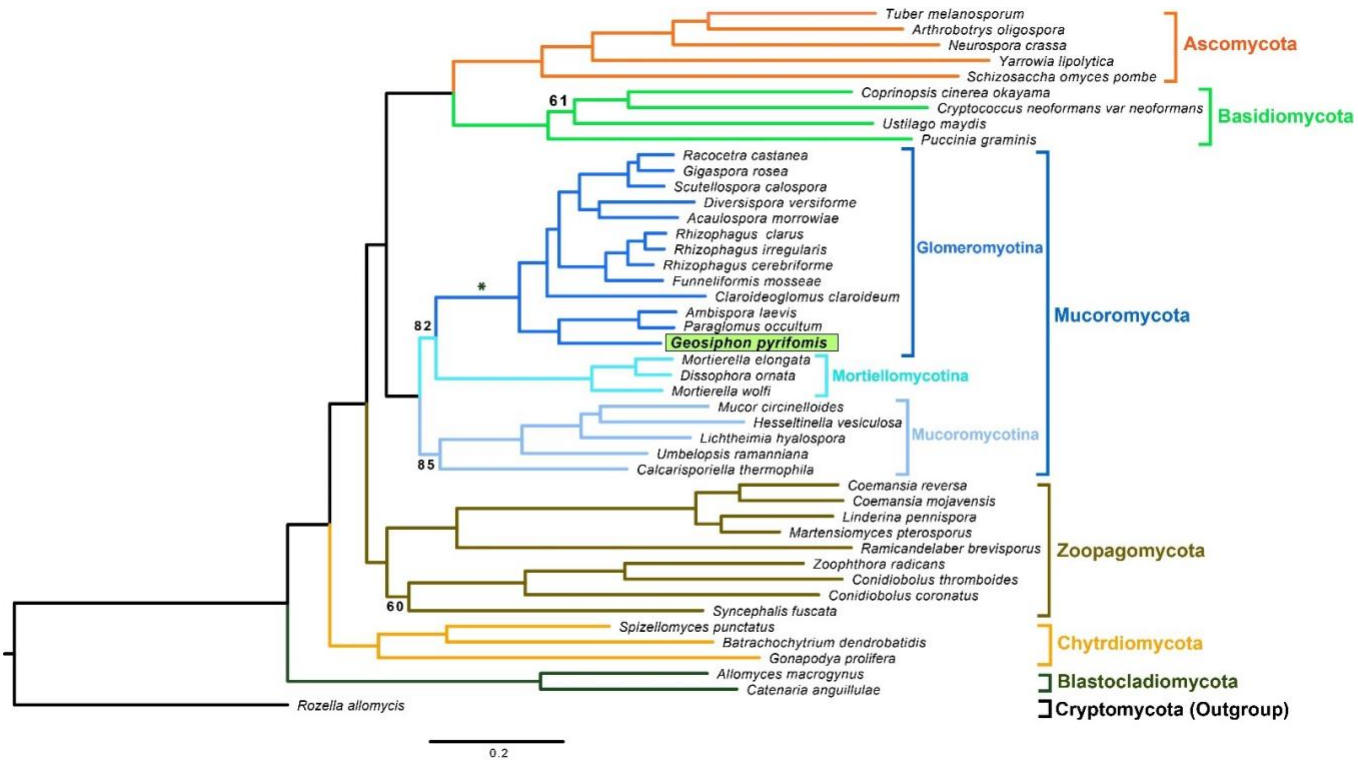
Using genome data and single nucleus data it was recently shown that AMF carry two genome organizations – i.e. homokaryotic (co-existing nuclei carry one parental haploid genotype) or dikaryotic (two parental genotypes co-exist in the mycelium) [26–29]. Mapping reads onto the *G. pyriformis* genome revealed reduced levels of polymorphism (0.5 SNP/Kb) and allelic frequencies suggesting that this species is carries low a low nuclear diversity and likely homokaryotic (Supplementary Figure 3).

Placement of G. pyriformis based on phylogenomics

The *G. pyriformis* genome annotation was used to identify the phylogenetic placement of this species using amino-acid sequences. In this case, we used a set of 434 conserved fungal single copy genes (data available at DOI: 10.5281/zenodo.1413687) to construct a phylogenetic tree of the fungal kingdom. Phylogenomics supports the monophyly of Glomeromycotina and its close

relationship with Mortierellomycotina within the phylum Mucoromycota [23] (Figure 3). Within Glomeromycotina, *G. pyriformis* groups within a monophyletic clade with *Ambispora leptoticha* and *Paraglomus occultum*, which diverged around 287 MYA (Supplementary Figure 4). This clade is distinct from more diverged nodes that contain sequenced representatives from Glomerales and Diversisporales [22]. The current placement of *G. pyriformis* as a sister lineage to *A. leptoticha* and *P. occultum* has full statistical support, and is favored by 73% of the gene sequences we used. Alternative topologies that place, for example, *G. pyriformis* at a basal node to all AMF (Alt-T1; Supplemental Figure 5, Supplemental Table 6) or as being associated with Glomerales or Diversisporales (Alt-T2 and Alt-T3; Supplemental Figure 5, Supplemental Table 6) were all rejected significantly using statistical tests implemented in IQ-TREE, including the KH, SH, ELW, and AU tests (Supplemental Table 6).

Figure 3: Phylogenetic tree representing the evolutionary relationships of fungi and placement



of *G. pyriformis* in Glomeromycotina clade. The tree was resolved using maximum likelihood phylogenetic reconstruction with IQ-TREE on a concatenated alignment of 434 protein coding genes. Numbers indicate nodes with less than 100% bootstrap support. Branches are coloured according to their phylum. Phylogenetic tests for all alternative topological placements of *G.*

pyriformis were rejected. The asterisk denotes the location where inferred gene losses and gains occurred in the MRCA of Glomeromycotina.

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

Phylogenomics revealed that *G. pyriformis* is a member of a clade that diverged early from the lineage encompassing the already sampled Diversisporales and Glomerales. As such, the *G. pyriformis* genome fills the gap in the genomic coverage of major AMF phylogenetic clades. With this data in hand, we first searched for genetic features that arose in the most recent common ancestors (MRCA) of all Glomeromycotina by comparing orthogroups from five available AMF genomes and four other members of the Mucoromycota as outgroups using OrthoFinder [39]. This analysis identified 661 gains and 344 losses that occurred before the divergence of the Glomeromycotina (Supplemental Table 7 and 8).

Among the 344 orthogroups classified as lost in the MRCA of the Glomeromycotina, we note the missing key enzymes involved in essential metabolic functions, such as sugar and thiamine metabolisms, or in the biosynthesis of fatty acids. These genes are also referred to as "Missing Glomeromycotina Core Genes" (MGCGs; Supplementary Table 8, Supplementary Table 9), and our analysis reveals that these have also been lost by *G. pyriformis*. Other key losses that affect all sequenced Glomeromycotina include enzymes that actively degrade plant cell wall (Supplementary Table 10). Among the 661 orthogroups gained in the MRCA, most encode for proteins involved in signaling pathways (e.g. protein kinases), protein–protein interactions (e.g. the tetratricopeptide repeat, Sel1, the homodimerization BTB (Broad-Complex, Tramtrack and Bric a brac), and WD-40 domain-containing proteins) and High Mobility Box (HMG) (Supplementary Table 11).

Comparative genomics also showed that *G. pyriformis* carries the same signatures of sexual reproduction found in AMF relatives. These include a complete set of meiosis-specific genes (Supplementary Table 12) [32], and a highly conserved genomic locus with architecture and sequence similarity to the mating-type (MAT) locus of basidiomycetes [33,34] (Supplementary Figure 6).

Regulation of orthogroups gained in the MRCA of Glomeromycotina

We investigated available gene expression data from the model AMF *Rhizophagus irregularis,* and found that 7 and 39 orthogroups (with a total of 8 and 272 genes; Supplementary Table 13 and Supplementary Table 14) gained in the MRCA of *Glomeromycotina* are respectively upregulated and downregulated across all four experimental conditions in the model AMF *Rhizophagus irregularis*.

These conditions include symbiotic associations between *R. irregularis* and distinct plant hosts (*Medicago truncatula, Brachypodium distachyon*) [7,43], and others based on laser-capture microdissection arbuscule-specific gene expression[34,35,37,39]. We investigated the putative function of these differentially regulated genes by identifying protein motifs along their coding sequences, and found that these are involved in a myriad putative cellular function, though these mostly include protein tyrosine kinases, cytochrome p450, as well as FAD binding (Supplemental Table 13 and 14). Lastly, one differentially regulated OG (OG0001728) shows evidence of originating from horizontal gene transfer from bacteria – i.e. this orthogroup is shared between bacteria and fungi (Supplementary Figure 7).

Distinct genomic features of Geosiphon pyriformis

Besides the evolution of AMS, the acquisition of the *G. pyriformis* genome offers an opportunity to identify innovations linked to the emergence of the only known cyanobacteria – fungus endosymbiosis. To identify such innovations, hierarchical clustering and abundance of Pfam domains was performed using available genomes in the Glomeromycotina and representatives of Mucoromycotina and Mortierellomycotina (Supplementary Figure 8). This analysis revealed a significant overrepresentation of 16 protein domains in *G. pyriformis* compared to relatives in the Mucoromycota – e.g. Lipase_3, RNase_H, Retrotrans gag domains, dUTPase, Spuma_A9PTase, Myb_DNA-bind_6 (Supplementary Table 15; Supplementary Table 16).

We also sought evidence of horizontal gene transfers (HGT) between partners of the unique *Geosiphon-Nostoc* endosymbiosis, and found 18 genes with potential bacterial within in the *G. pyriformis* genome (Supplementary Table 17). Among putative HGT, two are protein encoding genes with significant sequence conservation with *Nostoc* and Gamma proteobacteria homologues (Supplementary Figure 9, Supplementary Figure 10). All putative HGT are located

within contigs with average coverage and surrounded by genes of AMF origin, suggesting these do not represent contaminants.

Discussion

MRCA of all extant Glomeromycotina carried hallmarks of mutualism and obligate biotrophy

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

The retention of a conserved Glomeromycotina gene set in *G. pyriformis*, including a sub-set of these involved in plant cell wall degradation, is surprising given that this species was never seen producing mycorrhizae. As such, this retention could reflect an intrinsic capacity for *G. pyriformis* to undergo classic (but rare) mycorrhizal associations with plants under the right conditions. Although speculative at this point, this hypothesis is supported by the identification of rare *Geosiphon*-like sequences in environmental samples [26–28,36].

Novel symbiotic abilities and horizontal gene transfers in G. pyriformis

As a result of losses in fatty acid biosynthesis genes, AMF are entirely dependent on the host plant they associate with to obtain lipids [45,47,49–51]. Within this context, our findings suggest that, during the switch from regular AMS to a fungal- cyanobacteria symbiosis, *G. pyriformis* has evolved novel strategies to obtain lipids from its new host through the expansion of specific gene motifs. Specifically, the *G. pyriformis* genome carries a striking over-representation of Lipase 3 protein domains that hydrolyze ester linkages of fatty acids. As *Nostoc* spp is known to produce a wide variety of extracellular lipids in high amounts [52,53], it is possible that these abundant lipids are released in the environment (like many other cellular compounds released by cyanobacteria

[43,45,46,48]) and are then broken down by lipases to be used as an energy resource by *G. pyriformis*.

As we find evidence of bacteria-like genes in the *G. pyriformis* genome, our work also suggests that the co-existence of multiple endosymbionts and *G. pyriformis* nuclei within restricted bladders offers some opportunities for horizontal gene exchange. Although none of the putative bacterial genes we identified in *G. pyriformis* are functionally related, there is evidence that one is differentially regulated during AMF symbiosis, supporting the notion that bacteria-like genes can play a major role in fungal evolution [58].

Identification of a putative AMF core-symbiotic toolkit

The *G. pyriformis* genome also enabled the identification of a putative core AMF symbiotic toolkit conserved in all the sampled Glomeromycotina. This set of genes is differentially regulated in model AMF during symbiotic interactions with different plant hosts, including dicots, monocots and non-vascular plants, and thus provides a basis for future research on symbiosis-related mechanisms in these plant symbionts. The identification of a core set of gene gains specifically regulated during mycorrhizal symbiosis, and their conservation across the Glomeromycotina phylogeny, also provides support for the early emergence of symbiosis-specific gene functions in AMF over 400 million years ago, contemporaneously with the evolution of the first land plants [6,55,56]. Lastly as genetic transformation is currently unfeasible in Glomeromycotina, only assumptions can be proposed for the function of these putative core genes. However, as some encode for chitin synthases, one attractive hypothesis could be that some evolved for the production of short-chain chitooligosaccharides or lipo-chitololigosaccharides that are known symbiotic signals triggering the activation of the symbiotic program on the host plant [46,48].

Materials and Methods

Cultivation of *G. pyriformis* samples from natural habitat

G. pyriformis was sampled during autumn in the only known stable habitat near the village Bieber in the Spessart (Germany). Active bladders of the *Geosiphon-Nostoc* endosymbiosis were found in slightly acidic soil (pH 5). The bladders occurred close to the hornwort *Anthoceros* spp. and the liverwort *Blasia pusilla L.*, as these plants harbor the cyanobacteria needed to trigger the

Geosiphon-Nostoc endosymbiosis. After sampling spores and bladders were transferred to the institute in Průhonice and cultured in beakers [57], which contain a small pot with a sterile mixture of sand and soil (from the original habitat). The cultures were grown in a climate chamber at 18°C with 14 h light and 10 h night. The substrate is kept wet by a filter paper, which reaches from the substrate into a water reservoir in the beaker. To be maintained over time, cyanobacteria be frequently added to the cultures. For our cultures, *Nostoc punctiformis* was obtained from the Culture Collection of Algae (SAG) at the University of Göttingen (Germany) as strain SAG69.79 [60] [59].

Genome and transcriptome sequencing and assembly

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

Poor quality and adapter sequences were trimmed using Trimmomatic [61] with parameters the following parameters of ILLUMINACLIP:2:30:10 SLIDINGWINDOW:5:20 LEADING:5 TRAILING:5 MINLEN:50. The resulting 1 GB of non-redundant metagenome reads were assembled using metaSPAdes V3.12.0 [62]. Assembled contigs were binned on the basis of tetra nucleotide signature using CONCOCT [63], following part of the procedure used to assembly the genome of *Diversispora epigaea* [64]. Binned clusters were annotated using BLAST v 2.6.0+ [65] and clusters containing bacterial hits were removed. Using this approach, 21 bona-fide AMF clusters were retained, and used as reference to filter original paired-end and mate pair reads with BLAT v. 36x1 [66]. The reads which had mapped to the filtered contigs and matched by BLAT were then extracted to build cleaned sequence libraries that were assembled with MaSuRCA 3.3.0 [67]. Additional round of nr BLAST searches on MaSuRCA assembled contigs were performed to further remove contaminating bacteria. K-mer (k = 21) based methods were used on filtered

reads to estimate genome size of *G. pyriformis* using jellyfish 1.1.12 and plotted in GenomeScope 2.0 [68,69].

Genome annotation

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

Transposable elements were predicted using TransposonPSI [81]. Repeat sequences were first identified using RepeatModeler with multiple numbers of iterations. The iteration with the most number of repeats were then used for soft-masking the genome with REPEATMASKER (open 4.0.646) [82]. Output files generated from above procedures were used to identify repeat along the assembly. The completeness of genome assembly was assessed with BUSCO version 2.0 [83] with default parameters using the fungal gene dataset [fungi_odb9] [http://buscodev.ezlab.org/datasets/fungiodb9.tar.gz] with HMMER version 3.2.1 searches [84].

Putative gene functions were identified using Diamond BLASTX [85]. Pfam domain analysis were performed using hmmscan [76] and Carbohydrate-active enzymes (CAZYme) were identified using the dbCAN CAZy database [88]. Putatative CAZYmes were further verified through comparisons of data from Morin et al 2019 [21]. Secretory proteins were identified using previously published pipelines [21,89], and effectors were identified using EffectorP 2.0 [90]. The putative MAT loci of *Paraglomus* sp., and *R. irregularis* were identified by BLAST search procedures. The tests for a two-speed genome analysis was performed in part by measuring intergenic distance among genes in genome using a R script (code available in GitHub repository) [30]. For this study, published genomes of additional Glomeromycotina, Mortierellomycotina, and Mucoromycotina were downloaded from JGI portal MycoCosm database [DOI: 10.1093/nar/gkt1183].

SNP calling and allele frequency analysis

Filtered Mate Pair and Paired End reads were mapped onto the assembled *G. pyriformis* genome using the BWA-MEM v 0.7.17 algorithm [91] and sorted into a BAM file using samtools (v 1.9) [92]. Variants were called using FREEBAYES v1.2.0 [93] and filtered using vcftools [94].

Filtering cutoffs and procedures were as described by in Ropars et al. 2016 and Morin et al. 2019 [21,33]. Quality filtered variants and SNPs which passed filtering were used for constructing allele frequency plot using a custom R script (code available in GitHub repository).

Phylogenetic analysis and molecular dating

The phylogenomic analyses employed a set of 434 generally conserved and single-copy proteins in fungi (data available at DOI: 10.5281/zenodo.1413687), which were developed through efforts of the 1000 fungal genomes project and provided in the Joint Genome Institute MycoCosm site [1,95,96]. Profile-Hidden-Markov-Models of these markers were searched in the *Geosiphon* predicted protein sequences using HMMER3 (v3.1b2) and recovered 393 homologs (out of the 434) in total. The 434 markers in 45 included fungal genomes were further collapsed into 57 partitions using a greedy search embedded in PartitionFinder v.2.1.1 for consistent phylogenetic signals [97]. Phylogenetic trees were produced using the PHYling pipeline (data available DOI: 10.5281/zenodo.1257002) and with maximum likelihood method implemented in IQ-TREE (v.1.7-beta9) [98]. Concordance factors across the tree were calculated using the package implemented in IQ-TREE.

The divergence time of *Geosiphon* sp. from the clade of "*Ambispora leptoticha* and *Paraglomus occultum*" was estimated using the R8S v1.81 [99] with the phylogenetic tree reconstructed from the earlier step. We employed five calibration constraints to calibrate the tree, including the crown groups of Fungi (1100 MYA) [84], Dikarya (772 MYA) [84], Chytridiomycota (>573 MYA) [85–87], the MRCA of Chytridiomycota and terrestrial fungi (>750 MYA) [85–87], and Glomeromycotina (>460 MYA) [88]. The divergence time of each clade was inferred using the Langley-Fitch method with Powell algorithm [89–91].

Alternative topology test and dating analyses

To test the likelihood of other possible phylogenetic placements of *G. pyriformis*, we first reconstructed the associated phylogenetic trees using constraint tree topology as illustrated in Supplementary Figure 5 via "-g" option of the IQTREE package (iqtree-1.7-beta9) [92]. We then compared our best tree (shown as Figure 3) with alternative topologies to compute the log-likelihoods of the trees using Kishino-Hasegawa test, Shimodaira-Hasegawa test, expected likelihood weight, and approximately unbiased test via "-zb" and "-au" parameters in IQTREE [93–97]. All tests were performed with 10,000 resampling estimated log-likelihood (RELL)

method for reliable results. The best-fit substitution models for the genome-scale data matrix were estimated using ModelFinder implemented in IQTREE package [98].

Detection of putative horizontal gene transfers

To identify genes in *Geosiphon* that have potential origin in cyanobacteria, we compared the *Geosiphon* sp. genome to the available fungal and cyanobacterial genomes. To highlight potential HGT genes, we used a Python script (available in github repository) [105] to filter out genome component in *G. pyriformis* with higher similarity score to cyanobacteria than any fungi, excluding the *G. pyriformis* itself.

Gene orthology and evolution of symbiotic specific genes

Orthogroups resulting from the OrthoFinder run were parsed using a custom Python script. To be retained, an orthogroup had to fill the following conditions: any sequence from the non-AMS fungi, at least one sequence of *Geosiphon pyriformis*, one sequence of either *Rhizophagus irregularis* or *Rhizophagus cerebriforme* and at least one sequence of *Gigaspora rosea* or *Diversispora epigaea*. Reciprocally, orthogroups that could correspond to gene losses in the AMS fungi were extracted by retaining orthogroups with no sequences of AMS fungi and at least one sequence of each non-AMF fungi.

Orthogroups showing evidence of regulation in symbiotic conditions in *R. irregularis* were subjected to Maximum Likelihood (ML) analysis to check for the absence of non-AMS species. First, proteins contained in orthogroups were searched against the nine proteomes of 9 distinct species using the BLASTp+ v2.9.0 [109] with default parameters and an e-value threshold fixed at 1e-05 (threshold was set to 1e-03 when no non-AMS species sequences were identified). Then, proteins were aligned using MUSCLE v3.8.31 [110] with default parameters and resulting alignment trimmed to remove positions with more than 80% of gaps using trimAl v1.4rev22 [111]. Prior to ML reconstruction, best fitting evolution model was tested using ModelFinder[98] and then ML analysis was performed using IQ-TREE v1.6.1 [112] with 10,000 replicates of SH-aLRT. Trees were visualized and annotated with the iTOL platform v5.5 [113].

After first round of phylogeny, orthogroups showing an AMF specific pattern were blasted against the full MycoCosm database (1565 proteomes, last accessed: 03/01/2020) to confirm the

AMF-specific pattern and a phylogenetic analysis was performed following the procedure described above.

Differential expression analysis and combination of expression data to orthogroups

Expression data of *Rhizophagus irregularis* in four conditions were used to select orthogroups containing gene significantly deregulated in symbiosis for further analysis. Paired-end reads were trimmed and fragments mapped onto Rhiir2_1 genome assembly of *R. irregularis* (https://genome.jgi.doe.gov/Rhiir2_1/). Stringent settings of mapping were used (similarity and length read mapping criteria at 98% and 95%, respectively). Genes differentially expressed (DEG) in planta compared to extraradical mycelium were identified after EdgeR [108] normalization with a false discovery rate (FDR) correction using CLC Genomic Workbench (Qiagen). We retained genes showing an expression > 2- or < -2-fold times in planta compared to extraradical hyphae (FDR ≤ 0.05). Sets of 2683, 2518 and 2410 DEG were found in *M. truncatula, B. distachyon* and *L. cruciata* respectively. Detailed information on the data are available at the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) portal (accession no GSE67926). The analysis performed on RNA-seq data from arbuscocytes in *M. truncatula* [44] presented 6359 DEG.

Code availability

All scripts used to analyze are archived in https://github.com/madhubioinfo/Geosiphon.

Data availability

Genome assembly is available in NCBI with accession number of JAAOMT000000000. Genome sequencing reads are submitted in SRA with accession number of SRR11466073, Bioproject PRJNA610605, Biosample SAMN14307302. RNA-seq reads are available in SRA with accession of SRR12018969, SRR12018968, SRR12018970.

Acknowledgements

We thank Vasilis Kokkoris and Allison MacLean for comments on an earlier version of the manuscript. N.C research is funded by the discovery program of the Natural Sciences and Engineering Research council (RGPIN-2020-05643) and the Discovery Accelerator Supplements program (RGPAS-2020-00033). N.C is a University of Ottawa Research chair in Microbial Genomics. M.K. and C.K. were funded by the Czech Sciences Foundation (GAČR) as Junior

grant with the project number GJ16-16406Y. This work was also supported by the Agence Nationale de la Recherche (ANR) grant EVOLSYM (ANR-17-CE20-0006-01) to P.M.D, by the Bill and Melinda Gates Foundation as Engineering the Nitrogen Symbiosis for Africa (OPP1172165) to PM. D and by Laboratoire de Recherche en Sciences Végétales (LRSV) laboratory, which belongs to the TULIP Laboratoire d'Excellence (ANR-10-LABX-41). J.E.S. is a CIFAR Fellow in the program Fungal Kingdom: Threats and Opportunities. Y.W. and J.E.S. were supported by US National Science Foundation grants DEB-1441715 and DEB-1557110.

References

1. Spatafora, J.W., Chang, Y., Benny, G.L., Lazarus, K., Smith, M.E., Berbee, M.L., Bonito, G., Corradi, N., Grigoriev, I., Gryganskyi, A., *et al.* (2016). A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. Mycologia *108*, 1028–1046.

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

3. Brundrett, M.C. (2009). Mycorrhizal associations and other means of nutrition of vascular plants: Understanding the global diversity of host plants by resolving conflicting information and developing reliable means of diagnosis. Plant Soil *320*, 37–77.

4. Smith, S., and Read, D. (2008). Mycorrhizal Symbiosis.

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

6. Strullu-Derrien, C., Selosse, M.A., Kenrick, P., and Martin, F.M. (2018). The origin and evolution of mycorrhizal symbioses: from palaeomycology to phylogenomics. New Phytol. *220*, 1012–1030.

7. Delaux, P.M., Radhakrishnan, G. V., Jayaraman, D., Cheema, J., Malbreil, M., Volkening, J.D., Sekimoto, H., Nishiyama, T., Melkonian, M., Pokorny, L., *et al.* (2015). Algal ancestor of land plants was preadapted for symbiosis. Proc. Natl. Acad. Sci. U. S. A. *112*, 13390–13395.

8. 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., *et al.* (2020). An ancestral signalling pathway is conserved in intracellular symbioses-forming plant lineages. Nat. Plants *6*, 280–289.

9. Beaudet, D., Chen, E.C.H., Mathieu, S., Yildirir, G., Ndikumana, S., Dalpé, Y., Séguin, S., Farinelli, L., Stajich, J.E., and Corradi, N. (2018). Ultra-low input transcriptomics reveal the spore functional content and phylogenetic affiliations of poorly studied arbuscular mycorrhizal fungi. DNA Res. *25*, 217–227.

10. Gehrig, H., Schüßler, A., and Kluge, M. (1996). Geosiphon pyriforme, a fungus forming endocytobiosis with Nostoc (cyanobacteria), is an ancestral member of the glomales: Evidence by SSU rRNA analysis. J. Mol. Evol. *43*, 71–81.

11. Redecker, D., Morton, J.B., and Bruns, T.D. (2000). Ancestral lineages of arbuscular mycorrhizal fungi (Glomales). Mol. Phylogenet. Evol. *14*, 276–284.

12. Schüßler, A., and Walker, C. (2011). 7 Evolution of the 'Plant-Symbiotic' Fungal Phylum, Glomeromycota. In Evolution of Fungi and Fungal-Like Organisms, pp. 163–185.

13. Schüßler, A., Schwarzott, D., and Walker, C. (2001). A new fungal phylum, the Glomeromycota: Phylogeny and evolution. Mycol. Res. *105*, 1413–1421.

14. Kluge, M. (2002). A fungus eats a cyanobacterium: The story of the Geosiphon pyriformis endocyanosis. In Biology and Environment, pp. 11–14.

15. Mollenhauer, D., and Mollenhauer, R. (1988). Geosiphon cultures ahead. Endocytobios Cell Res *5*, 69–73.

16. Sun, X., Chen, W., Ivanov, S., MacLean, A.M., Wight, H., Ramaraj, T., Mudge, J., Harrison, M.J., and Fei, Z. (2019). Genome and evolution of the arbuscular mycorrhizal fungus Diversispora epigaea (formerly Glomus versiforme) and its bacterial endosymbionts. New Phytol. *221*, 1556–1573.

17. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

18. Smit, A.F.A., Hubley, R., and Green, P. (2010). RepeatMasker Open-3.0. 1996-2010. Inst. Syst. Biol.

19. Morin, E., Miyauchi, S., San Clemente, H., Chen, E.C.H., Pelin, A., de la Providencia, I., Ndikumana, S., Beaudet, D., Hainaut, M., Drula, E., *et al.* (2019). Comparative genomics of Rhizophagus irregularis, R. cerebriforme, R. diaphanus and Gigaspora rosea highlights specific genetic features in Glomeromycotina. New Phytol. *222*, 1584–1598.

20. Maeda, T., Kobayashi, Y., Kameoka, H., Okuma, N., Takeda, N., Yamaguchi, K., Bino, T., Shigenobu, S., and Kawaguchi, M. (2018). Evidence of non-tandemly repeated rDNAs and their intragenomic heterogeneity in Rhizophagus irregularis. Commun. Biol. *1*.

21. Tisserant, E., Malbreil, M., Kuo, A., Kohler, A., Symeonidi, A., Balestrini, R., Charron, P., Duensing, N., Frei Dit Frey, N., Gianinazzi-Pearson, V., *et al.* (2013). Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. Proc. Natl. Acad. Sci. U. S. A. *110*, 20117–20122.

22. Mathieu, S., Cusant, L., Roux, C., and Corradi, N. (2018). Arbuscular mycorrhizal fungi: intraspecific diversity and pangenomes. New Phytol. *220*, 1129–1134.

23. Chen, E.C.H., Morin, E., Beaudet, D., Noel, J., Yildirir, G., Ndikumana, S., Charron, P., St-Onge, C., Giorgi, J., Krüger, M., *et al.* (2018). High intraspecific genome diversity in the model arbuscular mycorrhizal symbiont Rhizophagus irregularis. New Phytol. *220*, 1161–1171. Available at: http://doi.wiley.com/10.1111/nph.14989 [Accessed September 9, 2019].

24. Dong, S., Raffaele, S., and Kamoun, S. (2015). The two-speed genomes of filamentous pathogens: Waltz with plants. Curr. Opin. Genet. Dev. *35*, 57–65.

25. Mathu Malar, C., Yuzon, J.D., Das, S., Das, A., Panda, A., Ghosh, S., Tyler, B.M., Kasuga, T., and Tripathy, S. (2019). Haplotype-phased genome assembly of virulent phytophthora ramorum isolate ND886 facilitated by long-read sequencing reveals effector polymorphisms and copy number variation. Mol. Plant-Microbe Interact. *32*, 1047–1060.

26. Sheng, M., Chen, X., Zhang, X., Hamel, C., Cui, X., Chen, J., Chen, H., and Tang, M. (2017). Changes in arbuscular mycorrhizal fungal attributes along a chronosequence of black locust (Robinia pseudoacacia) plantations can be attributed to the plantation-induced

variation in soil properties. Sci. Total Environ.

27. Panneerselvam, P., Kumar, U., Senapati, A., Parameswaran, C., Anandan, A., Kumar, A., Jahan, A., Padhy, S.R., and Nayak, A.K. (2020). Influence of elevated CO2 on arbuscular mycorrhizal fungal community elucidated using Illumina MiSeq platform in sub-humid tropical paddy soil. Appl. Soil Ecol.

28. Berruti, A., Demasi, S., Lumini, E., Kobayashi, N., Scariot, V., and Bianciotto, V. (2017). Wild Camellia japonica specimens in the Shimane prefecture (Japan) host previously undescribed AMF diversity. Appl. Soil Ecol.

29. Rich, M.K., Nouri, E., Courty, P.E., and Reinhardt, D. (2017). Diet of Arbuscular Mycorrhizal Fungi: Bread and Butter? Trends Plant Sci. *22*, 652–660.

30. Requena, N., and Breuninger, M. (2004). The Old Arbuscular Mycorrhizal Symbiosis in the Light of the Molecular Era. In, pp. 323–356.

31. Krüger, M., Krüger, C., Walker, C., Stockinger, H., and Schüßler, A. (2012). Phylogenetic reference data for systematics and phylotaxonomy of arbuscular mycorrhizal fungi from phylum to species level. New Phytol. *193*, 970–984.

32. Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. *16*.

33. Kamel, L., Tang, N., Malbreil, M., San Clemente, H., Le Marquer, M., Roux, C., and dit Frey, N.F. (2017). The comparison of expressed candidate secreted proteins from two arbuscular mycorrhizal fungi unravels common and specific molecular tools to invade different host plants. Front. Plant Sci. *8*.

34. Zeng, T., Holmer, R., Hontelez, J., te Lintel-Hekkert, B., Marufu, L., de Zeeuw, T., Wu, F., Schijlen, E., Bisseling, T., and Limpens, E. (2018). Host- and stage-dependent secretome of the arbuscular mycorrhizal fungus Rhizophagus irregularis. Plant J. *94*, 411–425.

35. An, J., Zeng, T., Ji, C., de Graaf, S., Zheng, Z., Xiao, T.T., Deng, X., Xiao, S., Bisseling, T., Limpens, E., *et al.* (2019). A Medicago truncatula SWEET transporter implicated in arbuscule maintenance during arbuscular mycorrhizal symbiosis. New Phytol.

36. Krüger, C., Kohout, P., Janoušková, M., Püschel, D., Frouz, J., and Rydlová, J. (2017). Plant communities rather than soil properties structure arbuscular mycorrhizal fungal communities along primary succession on a mine spoil. Front. Microbiol.

37. Gaude, N., Bortfeld, S., Duensing, N., Lohse, M., and Krajinski, F. (2012). Arbuscule-containing and non-colonized cortical cells of mycorrhizal roots undergo extensive and specific reprogramming during arbuscular mycorrhizal development. Plant J. *69*, 510–528.

38. Luginbuehl, L.H., Menard, G.N., Kurup, S., Van Erp, H., Radhakrishnan, G. V., Breakspear, A., Oldroyd, G.E.D., and Eastmond, P.J. (2017). Fatty acids in arbuscular mycorrhizal fungi are synthesized by the host plant. Science (80-.). *356*, 1175–1178.

39. Fiorilli, V., Catoni, M., Miozzi, L., Novero, M., Accotto, G.P., and Lanfranco, L. (2009). Global and cell-type gene expression profiles in tomato plants colonized by an arbuscular mycorrhizal fungus. New Phytol.

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

41. Keymer, A., Pimprikar, P., Wewer, V., Huber, C., Brands, M., Bucerius, S.L., Delaux, P.M., Klingl, V., von Röpenack-Lahaye, E., Wang, T.L., *et al.* (2017). Lipid transfer from plants to arbuscular mycorrhiza fungi. Elife *6*.

42. Jiang, Y., Wang, W., Xie, Q., Liu, N., Liu, L., Wang, D., Zhang, X., Yang, C., Chen, X., Tang, D., *et al.* (2017). Plants transfer lipids to sustain colonization by mutualistic mycorrhizal and parasitic fungi. Science (80-.). *356*, 1172–1173.

43. Peramuna, A., and Summers, M.L. (2014). Composition and occurrence of lipid droplets in the cyanobacterium Nostoc punctiforme. Arch. Microbiol. *196*, 881–890.

44. Temina, M., Rezankova, H., Rezanka, T., and Dembitsky, V.M. (2007). Diversity of the fatty acids of the Nostoc species and their statistical analysis. Microbiol. Res. *162*, 308–321.

45. López-Rosales, A.R., Ancona-Canché, K., Chavarria-Hernandez, J.C., Barahona-Pérez, F., Toledano-Thompson, T., Garduño-Solórzano, G., López-Adrian, S., Canto-Canché, B., Polanco-Lugo, E., and Valdez-Ojeda, R. (2019). Fatty acids, hydrocarbons and terpenes of

nannochloropsis and nannochloris isolates with potential for biofuel production. Energies.

46. Steinhoff, F.S., Karlberg, M., Graeve, M., and Wulff, A. (2014). Cyanobacteria in Scandinavian coastal waters - A potential source for biofuels and fatty acids? Algal Res.

47. Fitzpatrick, D.A. (2012). Horizontal gene transfer in fungi. FEMS Microbiol. Lett. *329*, 1–8.

48. Patel, V.K., Sundaram, S., Patel, A.K., and Kalra, A. (2018). Characterization of Seven Species of Cyanobacteria for High-Quality Biomass Production. Arab. J. Sci. Eng.

49. Rensing, S.A. (2018). Great moments in evolution: the conquest of land by plants. Curr. Opin. Plant Biol. *42*, 49–54.

50. Morris, J.L., Puttick, M.N., Clark, J.W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Yang, Z., Schneider, H., and Donoghue, P.C.J. (2018). The timescale of early land plant evolution. Proc. Natl. Acad. Sci. U. S. A. *115*, E2274–E2283.

51. Genre, A., Chabaud, M., Balzergue, C., Puech-Pagès, V., Novero, M., Rey, T., Fournier, J., Rochange, S., Bécard, G., Bonfante, P., *et al.* (2013). Short-chain chitin oligomers from arbuscular mycorrhizal fungi trigger nuclear Ca2+ spiking in Medicago truncatula roots and their production is enhanced by strigolactone. New Phytol. *198*, 190–202.

52. Maillet, F., Poinsot, V., André, O., Puech-Pagés, V., Haouy, A., Gueunier, M., Cromer, L., Giraudet, D., Formey, D., Niebel, A., *et al.* (2011). Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. Nature *469*, 58–64.

53. Schüßler, A., and Wolf, E. (2005). Geosiphon pyriformis—a Glomeromycotan Soil Fungus Forming Endosymbiosis with Cyanobacteria. In In Vitro Culture of Mycorrhizas, pp. 271–289.

54. Schlösser, U.G. (1994). SAG - Sammlung von Algenkulturen at the University of Göttingen Catalogue of Strains 1994. Bot. Acta *107*, 113–186.

55. v. Wettstein, F. (1915). Geosiphon Fr. Wettst, eine neue, interessante Siphonee. Österreichische Bot. Zeitschrift *65*, 145–156.

56. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

57. Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017). MetaSPAdes: A new versatile metagenomic assembler. Genome Res. *27*, 824–834.

58. Alneberg, J., Bjarnason, B.S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F., and Quince, C. (2014). Binning metagenomic contigs by coverage and composition. Nat. Methods *11*, 1144–1146.

59. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

60. Kent, W.J. (2002). BLAT---The BLAST-Like Alignment Tool. Genome Res. *12*, 656–664.

61. Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S.L., and Yorke, J.A. (2013). The MaSuRCA genome assembler. Bioinformatics *29*, 2669–2677.

62. Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics *27*, 764–770.

63. Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., and Schatz, M.C. (2017). GenomeScope: Fast reference-free genome profiling from short reads. In Bioinformatics, pp. 2202–2204.

64. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Philip, D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Macmanes, M.D., *et al.* (2013). De novo transcript sequence recostruction from RNA-Seq: reference generation and analysis with Trinity. Nat. Protoc. *8*, 1–43.

65. Hass, B. (2010). TransposonPSI: An application of PSI-Blast to mine (retro-)transposon ORF homologies. Broad Institute, Cambridge, MA, USA.

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

67. Buchfink, B., Xie, C., and Huson, D.H. (2014). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60.

68. Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., Park, Y.M., Buso,

N., and Lopez, R. (2015). The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic Acids Res. *43*, W580–W584.

69. Hancock, J.M., Zvelebil, M.J., Hancock, J.M., and Bishop, M.J. (2004). HMMer. In Dictionary of Bioinformatics and Computational Biology.

70. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. *42*.

71. Pellegrin, C., Morin, E., Martin, F.M., and Veneault-Fourrey, C. (2015). Comparative analysis of secretomes from ectomycorrhizal fungi with an emphasis on small-secreted proteins. Front. Microbiol. *6*.

72. Sperschneider, J., Dodds, P.N., Gardiner, D.M., Singh, K.B., and Taylor, J.M. (2018). Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. Mol. Plant Pathol. *19*, 2094–2110.

73. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

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

75. Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv Prepr. arXiv1207.3907. Available at: http://arxiv.org/abs/1207.3907.

76. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., *et al.* (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.

77. Ropars, J., Toro, K.S., Noel, J., Pelin, A., Charron, P., Farinelli, L., Marton, T., Krüger, M., Fuchs, J., Brachmann, A., *et al.* (2016). Evidence for the sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. Nat. Microbiol. *1*.

78. Spatafora, J.W., Chang, Y., Benny, G.L., Lazarus, K., Smith, M.E., Berbee, M.L., Bonito, G., Corradi, N., Grigoriev, I., Gryganskyi, A., *et al.* (2016). A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. Mycologia *108*, 1028–

1046.

79. Stajich, J.E. (2017). Fungal Genomes and Insights into the Evolution of the Kingdom. The Fungal Kingdom, 619–633.

80. Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., *et al.* (2014). MycoCosm portal: Gearing up for 1000 fungal genomes. Nucleic Acids Res. *42*.

81. Lanfear, R., Calcott, B., Ho, S.Y.W., and Guindon, S. (2012). PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. *29*, 1695–1701.

82. Cotet, G.B., Balgiu, B.A., and Negrea, V.C.Z. (2017). Assessment procedure for the soft skills requested by Industry 4.0. MATEC Web Conf. *121*.

83. Sanderson, M.J. (2003). r8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics *19*, 301–302.

84. Parfrey, L.W., Lahr, D.J.G., Knoll, A.H., and Katz, L.A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc. Natl. Acad. Sci. U. S. A. *108*, 13624–13629.

85. Lutzoni, F., Nowak, M.D., Alfaro, M.E., Reeb, V., Miadlikowska, J., Krug, M., Arnold, A.E., Lewis, L.A., Swofford, D., Hibbett, D., *et al.* (2018). Contemporaneous radiations of fungi and plants linked to symbiosis. Nat. Commun. *9*, 1–11.

86. Wang, Y., White, M.M., and Moncalvo, J.-M. (2019). Diversification of the gut fungi *Smittium* and allies (Harpellales) co-occurred with the origin of complete metamorphosis of their symbiotic insect hosts (lower Diptera). Mol. Phylogenet. Evol. *139*, 106550.

87. Chang, Y., Wang, S., Sekimoto, S., Aerts, A., Choi, C., Clum, A., LaButti, K., Lindquist, E., Ngan, C.Y., Ohm, R.A., *et al.* (2015). Phylogenomic analyses indicate that early fungi evolved digesting cell walls of algal ancestors of land plants. Genome Biol. Evol. *7*, 1590–1601.

88. Redecker, D. (2000). Glomalean Fungi from the Ordovician. Science *289*, 1920–1921.

89. Langley, C.H., and Fitch, W.M. (1974). An examination of the constancy of the rate of

molecular evolution. J. Mol. Evol. *3*, 161–177.

90. Gill, P.E., Murray, W., and Wright, M.H. (1981). Practical optimization (New York: Academic Press).

91. Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. (1992). Numerical recipes in C 2nd ed. (New York: Cambridge University Press).

92. Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.

93. Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. Syst. Biol. *51*, 492–508.

94. Shimodaira, H., and Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. *16*, 1114–1116.

95. Kishino, H., Miyata, T., and Hasegawa, M. (1990). Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J. Mol. Evol. *31*, 151–160.

96. Kishino, H., and Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence …. J. Mol. Evol. *29*, 170–179.

97. Strimmer, K., and Rambaut, A. (2002). Inferring confidence sets of possibly misspecified gene trees. Proc. R. Soc. B Biol. Sci. *269*, 137–142.

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

99. Wang, Y., White, M.M., Kvist, S., and Moncalvo, J.M. (2016). Genome-Wide Survey of Gut Fungi (Harpellales) Reveals the First Horizontally Transferred Ubiquitin Gene from a Mosquito Host. Mol. Biol. Evol. *33*, 2544–2554.

100. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403–410.

101. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792–1797.

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

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

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

Table 1: Summary statistics for genome assembly of sequenced *G. pyriformis* and other 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 %
Geosiphon Pyriformis	129	795	703	2733.91	0.023	64.35	96.2	29.25
Gigaspora rosea V1.0	597.95	7526	734	1204.75	7.92	63.44	97.9	28.81
Rhizophagus ceribriforme DAOM227022 V1.0	136.89	2592	266	709.02	17.60	24.77	98.3	26.55
Rhizophagus irregularis DAOM 197198V2.0	136.80	1123	129	1375.86	5.06	26.38	98	27.53
Diversispora versiformis strain IT104	147	731	434	2010.39	0.061	43.6	98.2	25.1
Rhizopus microsporus ATCC11559 V1	25.97	131	8	2782.17	2.41	4.68	98.6	37.48

Mucor Circinelloides CBS 277.49 V2.0	36.59	26	4	6050.25	0.00	20.38	97.2	42.17
Phycomyces Blakesleeanus NRRL1555 V2.0	53.94	80	11	4452.46	1.06	9.74	96.9	35.78
Mortierella Elongata AG-77	49.86	473	31	1526.29	0.30	4.63	99.7	48.05

Supplementary Figure legends

Supplementary Figure 1: Plot showing genome size of assembled genome. The x-axis represents the frequency of the number of times a given k-mer is observed in filtered illumina data of *Geosiphon pyriformis*. The y-axis represents the total number of k-mers with a given frequency.

Supplementary Figure 2: Heatmap showing evidence for a "one-speed genome" in *Geosiphon Pyriformis*. Density plot of the intergenic distances on the 5′ end (*x* axis) against the 3′ end (*y* axis) for each gene on the genome. Distances for all genes are binned and showed in color scale using a blue to red. red indicating the largest number of genes in a bin. Putative effectors are in black dots and found in gene dense region. In case of two speed genome, genome is separated in two compartment gene dense (containing essential housekeeping genes) and gene sparse region (containing secretory or effector/virulence related genes). Putative effector candidates are found in gene sparse repeat rich region.

Supplementary Figure 3: Genome wide allele frequency of *G. Pyriformis* shows it is a homokaryon. Allele distribution is based on read counts of filtered bi-allelic SNPs, and overlapped by density curves (black). Blue vertical lines represent the 0.5 allele frequency.

Supplementary Figure 4: Divergence time estimation of the *Geosiphon* sp. Branches lengths are proportional to time with the scale shown at the bottom. Estimated ages are labeled at each node. The ultrametric tree was produced using the r8s v1.81 program with the maximum likelihood tree obtained from the IQTREE program as the input.

Supplementary Figure 5: Alternative phylogenetic placements of *Geosiphon* examined in the alternative topology tests. Three alternative topologies with the *Geosiphon* at the earliest diverged position within Glomeromycotina (Alt_T1), grouped with the *Rhizophagus* and *Funneliformis* clade (Alt_T2), and joined with the *Gigaspora* and *Acaulospora* clade (Alt_T3) were compared to the best tree produced in the present study given the genome-wide data set. The log-likelihoods of each of the four topologies were computed using the IQTREE package and the results including the Kishino-Hasegawa test, Shimodaira-Hasegawa test, expected likelihood weight, and approximately unbiased test were summarized in Supplementary Table 6.

Supplementary Figure 6: Transcriptional directions of the putative AMF mating-type locus – i.e. HD1-like and HD2 - and their flanking regions in *G. pyriformis* and *R. irregularis*. The putative MAT locus and two surrounding genes are displayed in figure. Based on sequences initially identified by Ropars *et al.* 2016.

Supplementary Figure 7: Maximum Likelihood phylogeny of the OG0001728 obtained using the model: VT+R8. Bacterial proteins from nr database are indicated with red branches, fungal proteins from JGI database are indicated in orange branches. Sequences in blue constitute the original orthogroups.

Supplementary Figure 8: Presence and abundance of the different Pfam domain-containing proteins in the five genomes of Glomeromycota including *G. pyriformis* and Mucoromycota species. The heat map represents Pfam domain counts in each of the genomes. Only the 100 most represented domains are shown here. For visualization, the pfam abundance values are transformed into *z*-scores, which measures the relative enrichment (red) and depletion (green) of

each motif. The data were visualized and clustered using MultiExperiment Viewer (http://www.tm4.org/mev.html).

Supplementary Figure 9: Maximum likelihood phylogeny using WAG model of selenium binding protein from *G. pyriformis* showing HGT with other prokaryotes and eukaryotic relatives. Bootstrap values are provided for each node.

Supplementary Figure 10: Maximum likelihood phylogeny using WAG model of Molybdenum cofactor career protein from *G. pyriformis* showing HGT with homologues from prokaryotic and eukaryotic relatives. Bootstrap values are provided for each node.