RESEARCH

Whole genome analyses based on single, field collected spores of the arbuscular mycorrhizal fungus *Funneliformis geosporum*

Shadi Eshghi Sahraei[1] · Marisol Sánchez-García[1,2] · Merce Montoliu-Nerin[1,3] · David Manyara[1] ·
Claudia Bergin[4] · Søren Rosendahl[5] · Anna Rosling[1]

Abstract

Arbuscular mycorrhizal (AM) fungi are ubiquitous mutualistic symbionts of most terrestrial plants and many complete their lifecycles underground. Whole genome analysis of AM fungi has long been restricted to species and strains that can be maintained under controlled conditions that facilitate collection of biological samples. There is some evidence suggesting that AM fungi can adapt to culture resulting in phenotypic and possibly also genotypic changes in the fungi. In this study, we used field isolated spores of AM fungi and identified them as *Funneliformis geosporum* based on morphology and phylogenetic analyses. We separately assembled the genomes of two representative spores using DNA sequences of 19 and 22 individually amplified nuclei. The genomes were compared with previously published data from other members of Glomeraceae including two strains of *F. mosseae*. No significant differences were observed among the species in terms of gene content, while the single nucleotide polymorphism density was higher in the strains of *F. geosporum* than in the strains of *F. mosseae*. In this study, we demonstrate that it is possible to sequence and assemble genomes from AM fungal spores sampled in the field, which opens up the possibility to include uncultured AM fungi in phylogenomic and comparative genomic analysis and to study genomic variation in natural populations of these important plant symbionts.

Keywords AM fungi · Single spore · Morphology · Phylogeny · rDNA · Single nucleus sequencing

Introduction

Arbuscular mycorrhizal (AM) fungi form symbiotic interactions with the vast majority of plant species thus forming integral parts of terrestrial ecosystems (Parniske 2008; Smith

✉ Shadi Eshghi Sahraei
  shadi.eshghi.sahraei@ebc.uu.se

✉ Anna Rosling
  Anna.Rosling@ebc.uu.se

1 Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden

2 Department of Forest Mycology and Plant Pathology, Uppsala Biocentre, Swedish University of Agricultural Sciences, Uppsala, Sweden

3 Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden

4 Microbial Single Cell Genomics Facility, Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

5 Department of Biology, University of Copenhagen, Copenhagen, Denmark

and Read 2010). Plants benefit from the interaction with AM fungi through improved mineral nutrient and water uptake as well as increased stress tolerance. On the other hand, the AM fungi are obligate symbionts and obtain all their carbon and energy from their plant partners (Smith and Read 2010). Despite generally broad host ranges with little to no host preference (Bonfante and Genre 2010), we know that the species composition of AM fungal communities has functional consequences for terrestrial ecosystems (Hoeksema et al. 2018; Koziol et al. 2018). However, it is important to remember that most experimental studies exploring the biology of symbiosis are limited to AM fungi that can be maintained in culture, either in pots with their host plants or in vitro with root-organ cultures. Most AM fungi complete their life cycles underground, and their multinucleate asexual spores often are the largest biological unit that can be collected for morphological characterization and sequence identification (Bonfante and Genre 2010). Their biology has constrained whole genome analysis of AM fungi, and many studies have focused only on those species and strains from which ample biological

material can be obtained (Fortin et al. 2002; Abdellatif et al. 2019).

As a consequence of cultivability, species and strains in the genus *Rhizophagus*, which can be grown in vitro in root-organ cultures, are among the genomically most well-studied AM fungi (Kobayashi et al. 2018; Lin et al. 2014; Morin et al. 2019; Tisserant et al. 2013; Yildirir et al. 2022). Genomes have also been successfully sequenced and assembled from representatives of the order Diversisporales, such as *Diversispora epigaea* which is an AM fungus that forms above-ground sporocarps from which genome assemblies can be generated using metagenomic approaches (Sun et al. 2019) as well as two species in Gigasporales (Morin et al. 2019; Venice et al. 2020). Data from these important contributions have together demonstrated systematic features that distinguish the AM fungal genomes from their closest sister lineages. Together with the cyanobacteria symbiotic fungus *Geosiphon pyriformis*, AM fungi appear to have lost much of the ability to break down plant cell walls, in addition to genes involved in the synthesis of fatty acids and thiamine metabolism (Kobayashi et al. 2018; Malar et al. 2021). Future studies may shed light on gains and losses of specific genes in AM fungal lineages as taxon sampling has dramatically increased with recent development of genome assembly workflows from single nucleus sequencing (Montoliu-Nerin et al. 2020) making genome assemblies from species across seven families of AM fungi available (Montoliu-Nerin et al. 2021).

Interestingly, meta-analyses of soil community composition suggest that the AM fungal taxa that can be maintained in cultures are overrepresented in disturbed ecosystems and that wild plants host fewer cultured taxa than domesticated plants (Ohsowski et al. 2014). These observations indicate that ruderal life history strategies likely are overrepresented among AM fungal taxa used in controlled experimental studies and whole genome analyses (Ohsowski et al. 2014). Some AM fungi can be grown in pot culture for extended periods of time. For example, the most well-sequenced strain of *R. irregularis* (Tisserant et al. 2013; Yildirir et al. 2022) has been maintained in laboratory conditions since the 80 s (Stockinger et al. 2009). Emerging evidence suggests that growth under controlled conditions influences AM strains by selecting phenotypes that perform well under those conditions (Kokkoris and Hart 2019; Kokkoris et al. 2019). This highlights the importance of expanding whole genome analyses of AM fungi to ecologically relevant species and strains.

High levels of intra-specific genomic and phenotypic variation have been reported for AM fungi, including examples from the genus *Funneliformis* (Munkvold et al. 2004). Based on genetic markers, variation also has been observed within morpho-species assigned to *Funneliformis*, collected from fallow and cultivated soils (Rosendahl and Matzen 2008).

The genomic structure of genetic variation in AM fungi remains well studied; however, only in *R. irregularis* for which population sampling (Koch et al. 2004) has allowed the discovery of mating type linked heterokaryosis (Ropars et al. 2016). Furthermore, chromosome level assemblies has revealed genome structures with compartments of different variability (Yildirir et al. 2022). The aim of this study was to test if whole genome assemblies suitable for phylogenomic analysis could be generated from individually amplified and sequenced nuclei isolated from AM fungal spores collected from the field. With these assemblies in hand, we also explored within-genus differences regarding gene content and density of single nucleotide polymorphisms (SNPs).

Methods and materials

A single spore genome sequencing workflow developed by Montoliu-Nerin et al. (2020) was applied to spores from a field sampled morphotype of an AM fungus. Two spores were selected for nuclei isolation, sorting, amplification, and sequencing to assemble reference genomes as described below. Assemblies were used for species identification confirmation with established genetic markers. We then performed phylogenomic and comparative genomic analyses, estimates of SNPs, and characterization of the mating type (MAT) locus.

Field sampling and morphotyping of spores

A *Potentilla* sp. specimen with attached soil was collected using a hand shovel, at the Kungsängen Nature Reserve, N59 50′, E17 40′ (Uppsala, Sweden), on September 17th, 2019. This protected grassland is managed by grazing and the site is most famous as home to the largest population of *Fritillaria meleagris* in Sweden. The sample was stored at + 4 °C until the next day when soil (almost 30 g) was blended with 0.5L of tap water. After three pulses of blending, the suspension was separated through a set of sieves with sequentially different pore sizes (1 mm, 500 µm, 200 µm, and 38 µm) and rinsed with tap water. Particles from the 200-µm and 38-µm sieves were collected and transferred to Falcon tubes containing 20 mL of 60% sucrose solution. The tubes were centrifuged for 1 min at 3000 rpm and the supernatant was passed through a 5-cm diameter sieve with 38-µm pore size and rinsed thoroughly with tap water. The contents of the sieve were transferred to a petri dish for morphological examination under a Model SMZ800 stereomicroscope (Nikon, Japan). A number of spores identified as belonging to the genus *Funneliformis* (morpho-species *F. geosporum*) were separated on a moist filter paper (Fig. S1). Some of the spores were examined under an Olympus BX41 microscope (Olympus

corporation, Tokyo, Japan), to confirm morphotype identification. Remaining spores were transferred to an Eppendorf tube and washed with distilled water.

Single nuclei sorting, amplification, and sequencing

Two spores of the *F. geosporum* were placed in separate 1.5-mL tubes with 150 μL of $1 \times$ PBS (hereafter referred to as *F. geosporum* (A) for field collected spore A and *F. geosporum* (B) for field collected spore B). Then, spores were crushed with a sterile pestle and stained with 1 μL of $200 \times$ SYBR Green I Nucleic Acid stain (Invitrogen™, Thermo Fisher Scientific, MA, USA). Nuclei sorting was performed in accordance with a previously established protocol (Montoliu-Nerin et al. 2020) in which the solution was immediately transferred to a 0.5-mL tube to allow the debris to settle during the 30–60 min of staining. The nuclear sorting was performed at the SciLifeLab Microbial Single Cell Genomics Facility with a MoFlo™ Astrios EQ sorter (Beckman Coulter, USA) as in Montoliu-Nerin et al. (2020, 2021). From each spore, individual nuclei were isolated into 48 wells of one 384-well plate. Plates were stored at -80 °C until whole genome amplification with the enzyme Phi29 via multiple displacement amplification (MDA) was done, under clean conditions, using the Repli-g Single Cell kit (Qiagen) in a 10-μL reaction volume. The nucleic acid stain SYTO 13 was added to the reaction in order to monitor the DNA amplification over time. We performed a PCR amplification of rDNA markers on a 1:10 dilution of all amplified nucleus samples, using fungal and bacterial specific primers and following the protocol in Montoliu-Nerin et al. (2020). For each of the two isolates, a total of 22–24 amplified nuclei samples that tested positive for fungi and negative for bacteria were selected for sequencing (Supplementary datafile 1). Library preparation and sequencing was done at the SNP and SEQ Technology Platform in Uppsala at the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory (SciLifeLab), using the TruSeq PCR free DNA library preparation kit (Illumina Inc.), followed by 150 cycles of paired-end sequencing using the NovaSeq6000 system and v1 sequencing chemistry (Illumina Inc.).

Genome assembly

A total of 22 and 19 amplified nuclei were successfully sequenced from field collected spores A and B, respectively (Table S1). Reads obtained from the amplified nuclei were combined and then normalized using bbnorm of BBMap v.38.08 (Bushnell 2014, 2015) with a set average depth of $100 \times$. Thereafter, SPAdes v.3.12.0 (Bankevich et al. 2012) was used to generate one genome assembly per spore. Assembly statistics were obtained with QUAST v.5.0.2 (Gurevich et al. 2013), and genome completeness was estimated using

BUSCO v.3.0.2b with the fungi_odb9 as lineage setting, and *rhizopus_oryzae* as species set (Simão et al. 2015). To confirm spore identification to genus, the extracted SSU region was combined with the taxon rich SSU alignment from Krüger et al. (2012) and SSU regions extracted from previous AM genomes sequenced by our team (Montoliu-Nerin et al. 2021). To further confirm identity of the spores, phylogenetic analyses were performed on informative genomic regions with available reference data, including the D2 region of rDNA large subunit (LSU), intron 1 from FOX2, and intron 4 from TOR2. These regions were extracted from the assembly using BLASTn v.2.11.0 (Camacho et al. 2009). These markers have been characterized and used in previous studies of genetic variation among AM fungi (Krüger et al. 2012) and within the genus *Funneliformis* (Stukenbrock and Rosendahl 2005). FOX2 encodes a protein involved in peroxidal β-oxidation (Requena et al. 1999), and TOR2 encodes a protein involved in cell cycle processes (Requena et al. 1999). To resolve species identity of the spores, the extracted sequences were aligned with known *Funneliformis* sequences from GenBank and from local sequence resources. The alignment was visually inspected in AliView (Larsson 2014) before reconstructing a maximum likelihood (ML) phylogeny with 1000 bootstrap replicates using IQ-TREE v.2.0 (Minh et al. 2020).

Searching for a putative MAT-locus

To investigate the presence of the putative MAT-locus in the *Funneliformis* reference genome assemblies, we used the MAT-locus nucleotide sequence of *R. irregularis* (GenBank KT946687) as query in a BLASTn search in all *Funneliformis* assemblies. Requiring an e-value of zero identified two contigs in each of the four *Funneliformis* genome assemblies. The hit contigs were aligned with the query sequence and visually inspected in Geneious 11.0.5 (Biomatters Ltd). Only one contig from each assembly contained a possible homologue for one of the flanking regions (choline transporter-like protein) of the MAT-locus as annotated in KT946687. These four contigs were trimmed to cover the putative choline transporter-like homologue and regions that aligned within species. With the aim to determine if the analyzed *Funneliformis* genome assemblies represented strains that were homokaryotic or heterokaryotic for the MAT-locus (Ropars et al. 2016), we generated single nucleus genome assemblies and queried these with their trimmed contig including the putative choline transporter-like homologue. All single nucleus samples were individually assembled following the workflow for single nuclei assembly as described in Montoliu-Nerin et al. (2020). From the single nuclei assemblies, all contigs with a BLASTn hit score of zero were then aligned to the query and visually inspected using Geneious.

Taxon sampling

Based on the phylogenetic placement of *F. geosporum* (field collected spores A and B) using different gene regions (SSU, D1 of LSU, TOR2, and FOX2) extracted from the assemblies (Figs. S2–S5), we performed phylogenomic analyses of Glomeraceae. For this, we included six previously published genome assemblies of taxa within the family. Two published *F. mosseae* were included, of which the strain UK2014 was initially listed as *F. caledonius* but has been reidentified as *F. mosseae* (W. Wheeler personal communication) in line with phylogenomic analyses (Montoliu-Nerin et al. 2021). The genus *Rhizophagus* was represented by two published assemblies of the *R. irregularis* strain DAOM197198 (Chen et al. 2018; Montoliu-Nerin et al. 2021) and one assembly of the *R. irregularis* strain DAOM229456 (Morin et al. 2019). The latter was previously published under the name *R. diaphanus* (Morin et al. 2019), but phylogenetic analyses suggest that this strain actually belongs to *R. irregularis* (Błaszkowski et al. 2018). The genus *Oehlia* is represented by *O. diaphana* strain DAOM227022. A genome assembly of this strain was previously included in phylogenomic analyses under the name *R. cerebriforme* (Malar et al. 2021; Montoliu-Nerin et al. 2020; Morin et al. 2019), but based on molecular and morphological data, this strain was shown to represent *O. diaphana* (Błaszkowski et al. 2018). In the current study, names of all included taxa were corrected and updated in accordance with the latest taxonomy for these lineages (Błaszkowski et al. 2018) (Table S2). Based on a previous phylogenomic analysis of AM fungi (Montoliu-Nerin et al. 2021), we used published genome assemblies representing three taxa from the Diversisporales, and *Claroideoglomus candidum* as outgroup (Table S2).

Gene prediction

Gene prediction was performed on all genome assemblies generated from single nucleus sequencing with an in-house fungal annotation pipeline v.4.0, available at https://bitbucket.org/scilifelab-lts/genemark_fungal_annotation/src/master/. The pipeline first uses RepeatModeler v.1.0.8 (Smit and Hubley 2015) to de novo predict repeat elements from each assembly before masking repeats from the genome assembly with RepeatMasker v.4.0.7 (Smit et al. 2015). The pipeline then uses MAKER v.3.01.1-beta (Cantarel et al. 2008) to align the protein sequences in UniProt/Swiss-Prot (Bateman et al. 2021; UniProt Consortium 2018) to the masked genome assembly. Then, it uses GeneMark-ES v.4.33-es (Ter-Hovhannisyan et al. 2008) to de novo predict protein coding genes from the masked genome assemblies, guided by the locations of the aligned proteins from UniProt/Swiss-Prot. A minimum contig size of 10 Mb was calculated for each assembly to be included in

the self-training process of the GeneMark algorithm. For published genome assemblies, the gene predictions were included as published (Chen et al. 2018; Morin et al. 2019; Montoliu-Nerin et al. 2021).

Functional annotation

Predicted genes from all included genomes were functionally annotated using FunAnnotate v.1.8.9 (Palmer and Stajich 2020). Four databases were used to annotate the gene content, including InterProScan (Jones et al. 2014) for functional domain annotation, MEROPS v.12 (Rawlings et al. 2014) for peptidases, SignalP v.4.1f (Nielsen 2017) for secreted proteins, and dbCAN v.9 (Yin et al. 2012) to identify carbohydrate activated enzymes also known as CAZymes. For each genome assembly, the Pfam domains were normalized by the total number of identified domains in each taxon using the "decostand" function in the vegan package v2.5–7 (Oksanen et al. 2020). Domain composition among included genomes was then visualized using non-metric multi-dimensional scaling (nMDS) based on Bray–Curtis dissimilarity. For the other datasets, predicted genes in each gene family were normalized to the total number of predicted genes in each assembly (Table S3).

Phylogenomic and comparative analyses

Based on the amino acid sequences of predicted genes, we identified single copy orthologs (SCOs) present in all taxa using OrthoFinder v2.5.2 with default settings (Emms and Kelly 2019). All SCOs were individually aligned using MAFFT v.7.407 (Katoh and Standley 2013) before trimming with trimAL v.1.4.1 using a gap threshold of 0.1 (Capella-Gutiérrez et al. 2009). The individual SCO alignments were concatenated into a supermatrix using the script geneSticher.py (Schluter 2016), which also produces a partition file. Phylogenetic inference was performed with IQ-TREE v.2.0 running 1000 bootstrap replicates. To evaluate evolutionary relationships on a coalesce theory framework, we used ASTRAL-III v.5.7.3 (Zhang et al. 2018). For this analysis, single gene trees were generated using IQ-TREE v.2.0 and the best-fitting model option (-MFP) with 100 bootstrap iterations (Minh et al. 2020).

SNP calling analyses

Paired-end reads for all nuclei from each of the four *Funneliformis* spore samples (two cultivated and two collected from the field) were mapped against their respective reference genome assemblies using the BWA-MEM tool of the Burrows-Wheeler aligner (BWA) v.0.7.17 (Li and Durbin 2009) package.

SAMtools v.1.14 (Li et al. 2009) was used to convert SAM files to BAM files which were subsequently marked for duplicates, sorted by coordinates, and their read groups replaced using Picard v.2.23.4 (http://broadinstitute.github.io/picard/). SNPs were then called in individual nuclei using the Genome Analysis Toolkit (GATK) v.4.2.5.0 (McKenna et al. 2010) package. The HaplotypeCaller tool was first used to generate general variant calling files (gvcf) for each nucleus with the following parameters: ploidy = 1; minimum-mapping-quality = 30; minimum base quality = 20; max-alternate-alleles = 1. The gvcf files were merged into one gvcf file using the CombineGVCFs tool and subsequently genotyped with the GenotypeGVCFs tool. SNPs were hard-filtered using using GATK's recommended thresholds: QD, < 2; FS, > 60; MQ, < 40; MQRankSum, < −12.5; ReadPosRankSum, < −8; QUAL, < 30; SOR, > 3.0. An additional filtering was applied where sites in nuclei with less than five reads and allele fractions of more than 0.1 but less than 0.9 were marked as missing data, and sites across nuclei with more than 33% missing data were removed using a customized script (SNP_filter.R) available in https://github.com/drowl001/SNP_analysis. SNP densities were calculated for each strain as the number of SNPs in the non-repeat portions of the genome divided by the total size of the non-repeat portion of the genome assembly.

Results

A *F. geosporum* morpho-species was identified among AM fungal spores extracted from soil associated with a *Potentilla* sp. specimen (Fig. S1), and two representative spores were selected for genome sequencing. Both genome assemblies were over 130 Mb, with approximately 20,000 predicted genes and an estimated BUSCO completeness of 93% (Table 1). The MAT-locus could not be confidently identified in the *Funneliformis* genome assemblies. We determined one of the flanking gene regions, a choline transporter-like homologue, in all the reference assemblies, but single nuclei genome assemblies were fragmented and only a few contigs including the putative choline transporter-like region were recovered. Thus, with the available data we could not determine if the strains are homokaryotic or heterokaryotic for the MAT-locus.

Extracted SSU regions from the two genome assemblies were identical, and phylogenetic analysis, including published SSU sequences, firmly placed the spores in the genus *Funneliformis*, but identification to species level could not be achieved from the SSU dataset (Fig. S2). The phylogenetic analyses using the rDNA LSU D1 region included six species in the genus *Funneliformis* and despite slightly different LSU sequences, both spores were strongly supported as *F. geosporum* (Fig. S3) as expected from the morphological observations. For TOR2 and FOX2, identical sequences were extracted from the two genome assemblies. TOR2 analysis provided enough resolution to separate *F. geosporum* from sister species in the genus (Fig. S4). Whereas the phylogenetic signal in FOX2 was less clear at the genus level, and did not resolve *F. geosporum* and *F. caledonius* as monophyletic lineages. It did however group the sequences from the field collected spores in *F. geosporum* (Fig. S5). Based on the combined results, we conclude that the sequenced spores from the field represent *F. geosporum*.

Here we present genome assemblies from the two spores of *F. geosporum* extracted directly from field soil. This is the first direct genome sequencing of AM fungal spores sampled from a natural habitat. In a phylogenomic analysis that included members of Glomerales and a total of 1205 SCOs present in all taxa, both genomes of *F. geosporum* (field collected spores A and B) were recovered in a well-supported clade together with two strains of *F. mosseae* (Fig. 1). Our analyses demonstrate that *Funneliformis* is the sister genus to *Rhizophagus* and *Oehlia*. The latter is represented by only one species, *O. diaphana*, but the status as a distinct genus is well supported by its separation from *Rhizophagus* on a long branch (Fig. 1), and by morphological synapomorphies (Błaszkowski et al. 2018).

Comparative genome analyses across the AM fungi included here demonstrate highly similar gene content among strains in the genus *Funneliformis*. Ordination analyses of the Pfam domains show that all *Funneliformis* strains cluster together separated from other taxa (Fig. 2). Curiously, the three genome assemblies of *R. irregularis*, including two of the strain DAOM197198, do not form a similarly

Table 1 Genome assembly statistics for two spores of *Funneliformis geosporum* (field collected spores A and B) collected from soil attached to a specimen of *Potentilla* sp. from the Kungsängen Nature Reserve, Uppsala, Sweden

Species (ID)	Size (Mb)	Number of contigs	N50	Largest contig (kb)	GC (%)	BUSCO (%)	Number of genes	Repeats (Mb)
F. geosporum (A)	134.3	27,829	10,561	93,301	26.71	C: 93.1 (S:92.8, D:0.3), F:1.4	20,100	68.22
F. geosporum (B)	132.9	28,058	10,159	168,938	26.74	C: 93.1 (S:93.1, D:0.0), F:1	20,745	65.51

C completeness includes both single copy (S) and duplicated (D) genes, *F* fragmented. Given in % out of 290 BUSCO genes
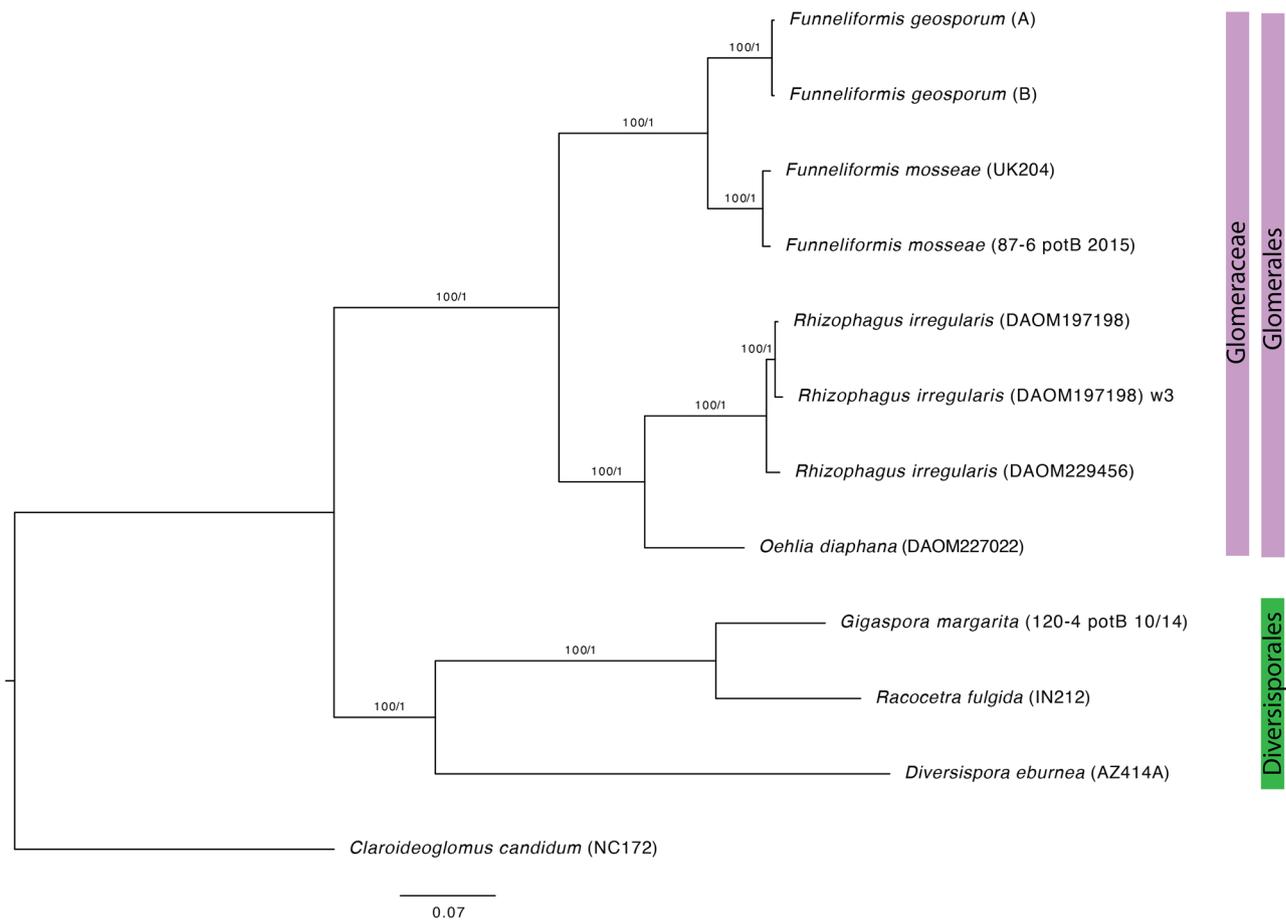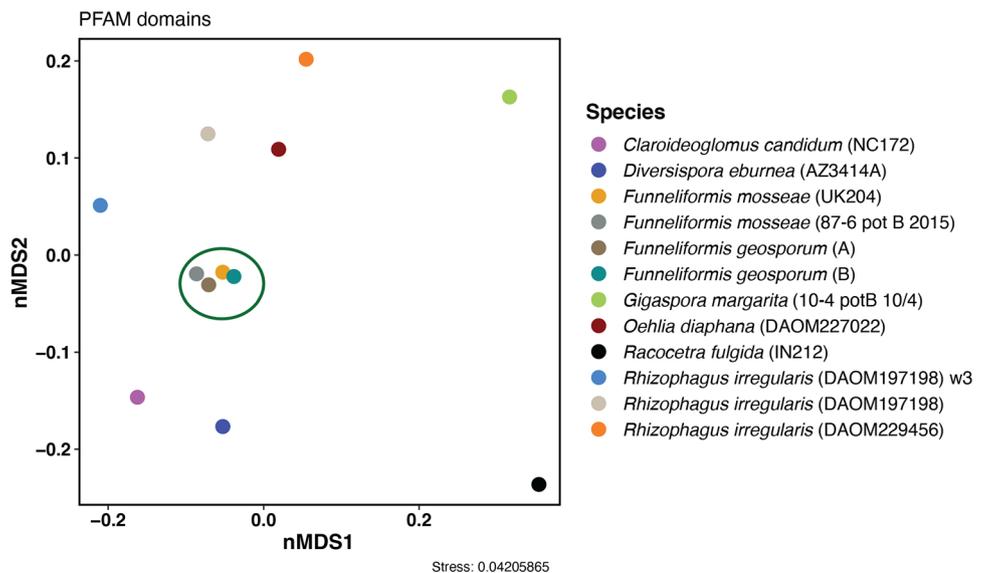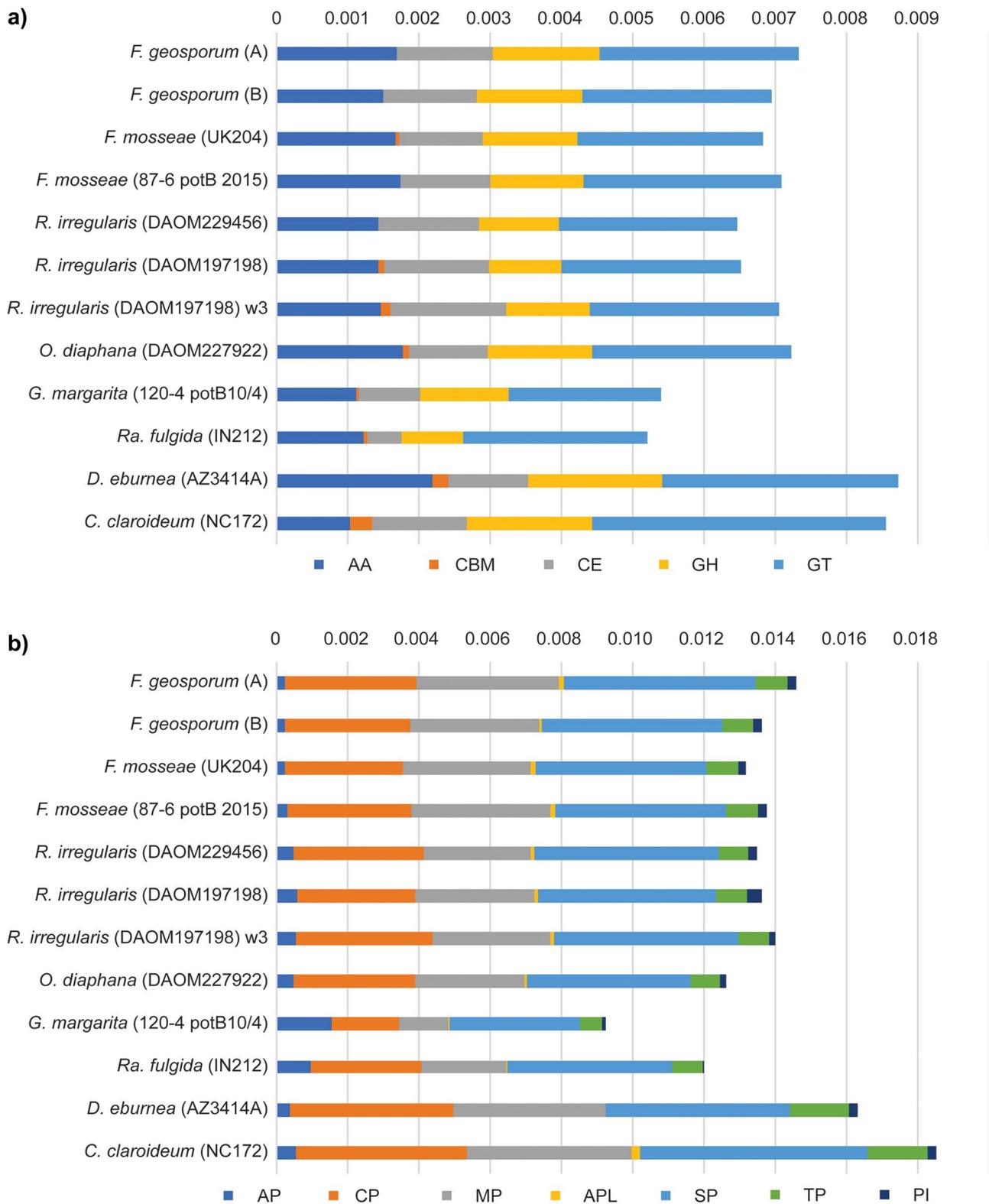
Fig. 1 Phylogeny of eight taxa from three genera in Glomerales, three taxa representing the order Diversisporales and *Claroideoglomus candidum* as outgroup. Best maximum likelihood IQ-TREE tree from a concatenated alignment of 1205 single copy orthologs (SCOs) shared among all taxa. The same topology was recovered in an ASTRAL analysis based on individual gene trees for the same SCOs. Multilocus bootstrapping and local posterior probabilities are indicated at the nodes (MLBS/LPP). All nodes have 100 bootstrap supports based on 100 replications and posterior probability of 1. *Funneliformis geosporum* A and B refer to field collected spores

Fig. 2 Non-metric multi-dimensional scaling (nMDS) based on functional domain annotations on relative abundance data of all Pfam domains using Bray–Curtis dissimilarity. *Funneliformis* taxa are highlighted with a green circle. *Funneliformis geosporum* A and B refer to field collected spores

Fig. 3 Genes across five carbohydrate-active enzyme (CAZyme) gene families (**a**) and seven peptidases gene families (MEROPS) (**b**), expressed as proportions of the total number of predicted genes from each genome assembly. CAZyme gene families in **a** include Auxiliary activities (AA), Carbohydrate binding modules (CBM), Carbohydrate esterase (CE), Glycoside hydrolase (GH), and Glycosyl transferases (GT). Note that no genes in the Polysaccharide lysases family was detected across the analyzed genome assemblies. The MEROPS gene families in **b** include Aspartic Peptidase (AP), Cysteine peptidase (CP), Metallo peptidase (MP), Asparagine peptide lyase (APL), Serine peptidase (SP), Threonine peptidase (TP), and Protease inhibitors (PT). *Funneliformis geosporum* A and B refer to field collected spores

tight cluster despite representing a single species (Fig. 2). An analysis of the number of genes across CAZyme gene families revealed high similarity in content across genomes in the genus *Funneliformis* (Fig. 3a, Table S4). None of the analyzed AM fungal genomes contained genes that belong to the Polysaccharide Lysase (PL) family. Few genes in the Carbohydrate Binding Module family (CBM) were found, with the highest record of five genes detected in the outgroup *C. candidum*. Across *Funneliformis*, only one of the *F. mosseae* strains (UK204) contained a gene annotated as member of the CBM family (Fig. 3a). Relative numbers of MEROPS gene families are highly comparable across *Funneliformis* and *Rhizophagus*. However, *Rhizophagus* had more genes in the Aspartic Peptidase (A) family (11–15 genes per genome assembly) compared to *Funneliformis* (5–6 genes per genome assembly) (Fig. 3b). The percentage of genes classified as secreted proteins varied across the analyzed AM fungi, from 0.4% in *G. margarita* to 3.8% in *Ra. fulgida* (Fig. S6). Corresponding numbers across strains in the genus *Funneliformis* were highly comparable with an average of 2.8%. Slightly higher values were obtained for *Rhizophagus* with an average of 3.6% of all genes being classified as secreted proteins (Fig. S6). None of the comparative genome analyses suggests any major differences in gene content composition for members of the genus *Funneliformis* whether isolated from field collected or cultured material.

All *Funneliformis* strains generally showed a low intraisolate genetic variation, with SNP densities ranging from 0.04 to 0.26 SNPs per kilobase in the non-repeat regions of the genome. The two *F. geosporum* strains show a higher within-strain variation with SNP densities at 0.24 and 0.26 SNPs/kb, compared to the two *F. mosseae* strains at 0.04 and 0.13 SNPs/kb respectively (Table 2).

Discussion

Our work demonstrates that genome assemblies can be generated from AM fungus spores collected in the field. The study was possible thanks to development of a workflow from single nuclei sorting through whole genome amplification and sequencing to a curated assembly pipeline (Montoliu-Nerin et al. 2020). Phylogenetic analysis of the two spores placed

them firmly in the genus *Funneliformis* as *F. geosporum* which is in accordance with the morphological characterization at the time of sampling. The number of predicted genes observed in the two genome assemblies from *F. geosporum* spores is very similar and of a quality well suited for phylogenomic analysis. Extraction of marker genes such as D1 of LSU, TOR2, and FOX2, followed by focused phylogenetic analysis with known references, proved to be an efficient method for species identification. While slightly different LSU sequences were extracted, the TOR2 and FOX2 sequence were identical for the two spores (Fig. S3–S5). This is probably because the rDNA operon occurs as multiple separate copies in AM fungi (Maeda et al. 2018) and different variants were likely assembled in this study. FOX2 and TOR2 on the other hand are single copy genes.

In addition to the field collected *F. geosporum* spores sequenced in this study, our analysis included two *F. mosseae* strains (Montoliu-Nerin et al. 2021). Of these, the strain *F. mosseae* 87–6 potB 10/4 has been in culture since 2003 while the strain *F. mosseae* UK204 has been in culture since 1994. Assembly statistics (Table 1, Table S3) and genome annotation comparisons (Fig. 2, S6) show no indication of differences in genome content between taxa harvested directly from the field compared to taxa maintained in culture for many generations. Nevertheless, it is possible that other aspects of genome organization not covered by gene annotations could differ between cultured and field collected strains. For instance, considerable chromosome length variation and structural rearrangements have been observed across three well studied strains of *R. irregularis* separately isolated from the same field (Yildirir et al. 2022). Furthermore, we acknowledge that an appropriate comparison of genetic variation would require that field collected and cultured strains of the same species are analyzed.

While all four *Funneliformis* strains cluster tightly in the ordination analysis of functional domains, the three assemblies of *R. irregularis* were far more dispersed in the two visualized dimensions (Fig. 2). This might be because of large variation in gene content among *Rhizophagus* strains, where large copy number variation has been reported (Mathieu et al. 2018). It also might be due to differences in the quality of the genome assemblies and annotations (Table S3). Genome assemblies generated from single amplified and sequenced nuclei are

Table 2 SNP statistics for two published strains of *F. mosseae* and two field collected spores (A and B) of *F. geosporum*

	F. mosseae UK204	*F. mosseae* 87–6 pot B 2015	*F. geosporum* (A)	*F. geosporum* (B)
Total number of SNPs	84,819	46,678	110,604	94,216
Number of SNPs non-rep	9808	2899	17,404	15,370
Number of SNPs coding region	2563	967	7184	5934
Density SNP/kb[*]	0.13	0.04	0.26	0.24

[*]Density calculated as the number of SNPs in the non-repeat region of the genome divided by the size of the non-repeat region of the genome assembly

notably fragmented, with a N50 of 21,000 kb compared to 51,000 kb (Chen et al. 2018) and 137,000 (Morin et al. 2019) for assemblies generated from high-quality DNA extracts. With decreasing fragmentation, more genes can be properly predicted and annotated potentially affecting the results of overall protein family analyses (Fig. 3). However, the *R. irregularis* assembly generated from amplified and sequenced nuclei has been shown to be highly similar to a conventionally sequenced genome of the same strain (Montoliu-Nerin et al. 2021), and it is likely that there is a strong biological signal in the observed differences in functional domain composition within the two genera in Glomeraceae.

Analysis of SNP density suggests a higher genetic diversity for the field collected spores of *F. geosporum* (0.24–0.26 SNPs/kb) than for the cultured strains (0.04–0.13 SNPs/kb) of *F. mosseae*. The numbers are comparable to published SNP density estimates from other cultured AM fungus strains. SNP densities around 0.2 SNPs/kb have been estimated in *R. irregularis* strains that are homokaryons for the MAT-locus, while strains that are heterokaryons for the MAT-locus have higher densities from 0.45 to 0.8 SNPs/kb (Ropars et al. 2016). Observed differences in SNP density between genome assemblies of *F. geosporum* and *F. mosseae* may reflect differences in genome organization with regard to the mating type of nuclei in the strains. In the current study, however, we could not determine if the analyzed genome assemblies represent strains that are homokaryotic or heterokaryotic for the MAT-locus. This will be possible to resolve in the future when the MAT-locus is identified in *Funneliformis*.

In conclusion, a single-nucleus sequencing approach to assemble the whole genome of spores directly obtained from the field opens the door for extensive genome sampling of hitherto un-sequenced diversity among AM fungi. For instance, genomes could be assembled for species that are abundantly detected in undisturbed natural settings but cannot be grown in culture (Ohsowski et al. 2014). Furthermore, the method could facilitate population genetic studies of natural populations by extracting marker genes from genome assemblies representing different spores or could enable estimating genetic variation within spores to refine our understanding of how geographic distance and agricultural practices affect the population structure of AM fungi (Rosendahl and Matzer 2008; Rosendahl et al. 2009).

Declarations

References

Abdellatif L, Lokuruge P, Hamel C (2019) Axenic growth of the arbuscular mycorrhizal fungus *Rhizophagus irregularis* and growth stimulation by coculture with plant growth-promoting rhizobacteria. Mycorrhiza 29(6):591–598. https://doi.org/10.1007/s00572-019-00924-z

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Av P, Av S, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5):455–477. https://doi.org/10.1089/cmb.2012.0021

Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Bursteinas B, Bye-A-Jee H, Coetzee R, Cukura A, Ada S, Denny P, Dogan T, Ebenezer TG, Fan J, Castro LG et al (2021) UniProt: the universal protein

knowledgebase in 2021. Nucleic Acids Res 49(D1):D480–D489. https://doi.org/10.1093/nar/gkaa1100

Błaszkowski J, Kozłowska A, Niezgoda P, Goto BT, Dalpé Y (2018) A new genus, *Oehlia* with *Oehlia diaphana* comb. Nov. and an emended description of *Rhizoglomus vesiculiferum* comb. nov. in the Glomeromycotina. Nova Hedwigia, 107(3–4):501–518. https://doi.org/10.1127/nova_hedwigia/2018/0488

Bonfante P, Genre A (2010) Mechanisms underlying beneficial plant - fungus interactions in mycorrhizal symbiosis. In Nature Communications 1(48):1–11. https://doi.org/10.1038/ncomms1046

Bushnell B (2014) BBMap short read aligner. In Joint Genome Institute, department of energy (Vol. 13)

Bushnell B (2015) BBMap (version 37.75). Available at https://Sourceforge.Net/Projects/Bbmap/

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC Bioinformatics 10(421):1–9. https://doi.org/10.1186/1471-2105-10-421

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Alvarado AS, Yandell M (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18(1):188–196. https://doi.org/10.1101/gr.6743907

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25(15):1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Chen ECH, Morin E, Beaudet D, Noel J, Yildirir G, Ndikumana S, Charron P, St-Onge C, Giorgi J, Krüger M, Marton T, Ropars J, Iv G, Hainaut M, Henrissat B, Roux C, Martin F, Corradi N (2018) High intraspecific genome diversity in the model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*. New Phytol 220(4):1161–1171. https://doi.org/10.1111/nph.14989

Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20(238):1–11. https://doi.org/10.1186/s13059-019-1832-y

Fortin JA, Bécard G, Declerck S, Dalpé Y, St-Arnaud M, Coughlan AP, Piché Y (2002) Arbuscular mycorrhiza on root-organ cultures. In Canadian Journal of Botany 80:1–20. https://doi.org/10.1139/b01-139

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29(8):1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hoeksema JD, Bever JD, Chakraborty S, Chaudhary VB, Gardes M, Gehring CA, Hart MM, Housworth EA, Kaonongbua W, Klironomos JN, Lajeunesse MJ, Meadow J, Milligan BG, Piculell BJ, Pringle A, Rúa MA, Umbanhowar J, Viechtbauer W, Wang YW, Wilson GWT, Zee PC (2018) Evolutionary history of plant hosts and fungal symbionts predicts the strength of mycorrhizal mutualism. Communications Biology 1(116):1–10. https://doi.org/10.1038/s42003-018-0120-9

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30(4):772–780. https://doi.org/10.1093/molbev/mst010

Kobayashi Y, Maeda T, Yamaguchi K, Kameoka H, Tanaka S, Ezawa T, Shigenobu S, Kawaguchi M (2018) The genome of *Rhizophagus clarus* HR1 reveals a common genetic basis for auxotrophy among arbuscular mycorrhizal fungi. BMC Genomics 19(465):1–11. https://doi.org/10.1186/s12864-018-4853-0

Koch AM, Kuhn G, Fontanillas P, Fumagalli L, Goudet J, Sanders IR (2004) High genetic variability and low local diversity in a population of arbuscular mycorrhizal fungi. Proc Natl Acad Sci USA 101(8):2369–2374. https://doi.org/10.1073/pnas.0306441101

Kokkoris V, Hart M (2019) In vitro propagation of arbuscular mycorrhizal fungi may drive fungal evolution. Front Microbiol 10(2420):1–9. https://doi.org/10.3389/fmicb.2019.02420

Kokkoris V, Miles T, Hart MM (2019) The role of in vitro cultivation on asymbiotic trait variation in a single species of arbuscular mycorrhizal fungus. Fungal Biol 123(4):307–317. https://doi.org/10.1016/j.funbio.2019.01.005

Koziol L, Schultz PA, House GL, Bauer JT, Middleton EL, Bever JD (2018) The plant microbiome and native plant restoration: the example of native mycorrhizal fungi. In BioScience 68(12):996–1006. https://doi.org/10.1093/biosci/biy125

Krüger M, Krüger C, Walker C, Stockinger H, Schüßler A (2012) Phylogenetic reference data for systematics and phylotaxonomy of arbuscular mycorrhizal fungi from phylum to species level. New Phytol 193(4):970–984. https://doi.org/10.1111/j.1469-8137.2011.03962.x

Larsson A (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30(22):3276–3278. https://doi.org/10.1093/bioinformatics/btu531

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25(14):1754–1760

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment / Map (SAM) format and SAMtools 1000 Genome Project Data Processing Subgroup. Bioinformatics (oxford, England) 25(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lin K, Limpens E, Zhang Z, Ivanov S, Saunders DGO, Mu D, Pang E, Cao H, Cha H, Lin T, Zhou Q, Shang Y, Li Y, Sharma T, van Velzen R, de Ruijter N, Aanen DK, Win J, Kamoun S, Bisseling T, Geurts R, Huang S (2014) Single nucleus genome sequencing reveals high similarity among nuclei of an endomycorrhizal fungus. PLoS Genet 10(1):e1004078. https://doi.org/10.1371/journal.pgen.1004078

Maeda T, Kobayashi Y, Kameoka H, Okuma N, Takeda N, Yamaguchi K, Bino T, Shigenobu S, Kawaguchi M (2018) Evidence of non-tandemly repeated rDNAs and their intragenomic heterogeneity in *Rhizophagus irregularis*. Communications Biology 1(87):1–13. https://doi.org/10.1038/s42003-018-0094-7

Malar M, Krüger M, Krüger C, Wang Y, Stajich JE, Keller J, Chen ECH, Yildirir G, Villeneuve-Laroche M, Roux C, Delaux PM, Corradi N (2021) The genome of *Geosiphon pyriformis* reveals ancestral traits linked to the emergence of the arbuscular mycorrhizal symbiosis. Curr Biol 31(7):1570–1577. https://doi.org/10.1016/j.cub.2021.01.058

Mathieu S, Cusant L, Roux C, Corradi N (2018) Arbuscular mycorrhizal fungi: intraspecific diversity and pangenomes. New Phytol 220(4):1129–1134. https://doi.org/10.1111/nph.15275

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20(9):1297–1303. https://doi.org/10.1101/gr.107524.110

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R, Teeling E (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37(5):1530–1534. https://doi.org/10.1093/molbev/msaa015

Montoliu-Nerin M, Sánchez-García M, Bergin C, Grabherr M, Ellis B, Kutschera VE, Kierczak M, Johannesson H, Rosling A (2020) Building de novo reference genome assemblies of complex eukaryotic microorganisms from single nuclei. Sci Rep 10(1303):1–10. https://doi.org/10.1038/s41598-020-58025-3

Montoliu-Nerin M, Sánchez-García M, Bergin C, Kutschera VE, Johannesson H, Bever JD, Rosling A (2021) In-depth phylogenomic analysis of arbuscular mycorrhizal fungi based on a comprehensive set of de novo genome assemblies. Frontiers in Fungal Biology 2(716385):1–13. https://doi.org/10.3389/ffunb.2021.716385

Morin E, Miyauchi S, San Clemente H, Chen ECH, Pelin A, de la Providencia I, Ndikumana S, Beaudet D, Hainaut M, Drula E, Kuo A, Tang N, Roy S, Viala J, Henrissat B, Grigoriev Iv, Corradi N, Roux C, Martin FM (2019) Comparative genomics of *Rhizophagus irregularis*, *R. cerebriforme*, *R. diaphanus* and *Gigaspora rosea* highlights specific genetic features in Glomeromycotina. New Phytol 222(3):1584–1598. https://doi.org/10.1111/nph.15687

Munkvold L, Kjøller R, Vestberg M, Rosendahl S, Jakobsen I (2004) High functional diversity within species of arbuscular mycorrhizal fungi. New Phytol 164(2):357–364. https://doi.org/10.1111/j.1469-8137.2004.01169.x

Nielsen H (2017) Predicting secretory proteins with SignalP. Methods Mol Biol 1611:59–73

Ohsowski BM, Zaitsoff PD, Öpik M, Hart MM (2014) Where the wild things are: looking for uncultured Glomeromycota. New Phytol 204(1):171–179. https://doi.org/10.1111/nph.12894

Oksanen J, Blanchet F, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin P, O'Hara R, Simpson G, Solymos P, Stevens M, Szoecs E, Wagner H (2020) Vegan: community ecology package. (version 2.5–7). https://CRAN.R-project.org/package=vegan

Palmer JM, Stajich J (2020) Funannotate V1.8.1: eukaryotic genome annotation. https://funannotate.readthedocs.io/en/latest/

Parniske M (2008) Arbuscular mycorrhiza: the mother of plant root endosymbioses. In Nature Reviews Microbiology 6(10):763–775. https://doi.org/10.1038/nrmicro1987

Rawlings ND, Waller M, Barrett AJ, Bateman A (2014) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res 42(D1):D343–D350. https://doi.org/10.1093/nar/gkt953

Requena N, Füller P, Franken P (1999) Molecular characterization of GmFOX2, an evolutionarily highly conserved gene from the mycorrhizal fungus *glomus mosseae*, down-regulated during interaction with rhizobacteria. Mol Plant Microbe Interact 12(10):934–942. https://doi.org/10.1094/MPMI.1999.12.10.934

Ropars J, Toro KS, Noel J, Pelin A, Charron P, Farinelli L, Marton T, Krüger M, Fuchs J, Brachmann A, Corradi N (2016) Evidence for the sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. Nat Microbiol 1(6):1–9. https://doi.org/10.1038/nmicrobiol.2016.33

Rosendahl S, Matzen HB (2008) Genetic structure of arbuscular mycorrhizal populations in fallow and cultivated soils. New Phytol 179(4):1154–1161. https://doi.org/10.1111/j.1469-8137.2008.02535.x

Rosendahl S, McGee P, Morton JB (2009) Lack of global population genetic differentiation in the arbuscular mycorrhizal fungus *Glomus mosseae* suggests a recent range expansion which may have coincided with the spread of agriculture. Mol Ecol 18(20):4316–4329. https://doi.org/10.1111/j.1365-294X.2009.04359.x

Stockinger H, Walker C, Schüßler A (2009) "*Glomus intraradices* DAOM197198", a model fungus in arbuscular mycorrhiza research, is not *Glomus intraradices*. New Phytol 183(4):1176–1187. https://doi.org/10.1111/j.1469-8137.2009.02874.x

Schluter D (2016) Phylogenetic tools phylogenetic trees. https://github.com/ballesterus/Utensils

Simão FA, Waterhouse RM, Ioannidis P, Ev K, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Smit A, Hubley R (2015) RepeatModeler open-1.0. In Available at http://www.repeatmasker.org

Smit A, Hubley R, Grenn P (2015) RepeatMasker Open-4.0. RepeatMasker Open

Smith S, Read DJ (2010) Mycorrhizal symbiosis, 3rd edn. Academic press, SanDiego, USA

Stukenbrock EH, Rosendahl S (2005) Development and amplification of multiple co-dominant genetic markers from single spores of arbuscular mycorrhizal fungi by nested multiplex PCR. Fungal Genet Biol 42(1):73–80. https://doi.org/10.1016/j.fgb.2004.10.004

Sun X, Chen W, Ivanov S, MacLean AM, Wight H, Ramaraj T, Mudge J, Harrison MJ, Fei Z (2019) Genome and evolution of the arbuscular mycorrhizal fungus *Diversispora epigaea* (formerly *Glomus versiforme*) and its bacterial endosymbionts. New Phytol 221(3):1556–1573. https://doi.org/10.1111/nph.15472

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M (2008) Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res 18(12):1979–1990. https://doi.org/10.1101/gr.081612.108

Tisserant E, Malbreil M, Kuo A, Kohler A, Symeonidi A, Balestrini R, Charron P, Duensing N, Frei Dit Frey N, Gianinazzi-Pearson V, Gilbert LB, Handa Y, Herr JR, Hijri M, Koul R, Kawaguchi M, Krajinski F, Lammers PJ, Masclaux FG et al (2013) Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. Proc Natl Acad Sci USA 110(50):20117–20122. https://doi.org/10.1073/pnas.1313452110

UniProt Consortium (2018) Erratum: UniProt: the universal protein knowledgebase Nucleic acids research (2017) Nucleic Acids Res 46(5):D158–D169. https://doi.org/10.1093/nar/gky092

Venice F, Ghignone S, Salvioli di Fossalunga A, Amselem J, Novero M, Xianan X, Sędzielewska Toro K, Morin E, Lipzen A, Iv G, Henrissat B, Martin FM, Bonfante P (2020) At the nexus of three kingdoms: the genome of the mycorrhizal fungus *Gigaspora margarita* provides insights into plant, endobacterial and fungal interactions. Environ Microbiol 22(1):122–141. https://doi.org/10.1111/1462-2920.14827

Yildirir G, Sperschneider J, Malar CM, Chen ECH, Iwasaki W, Cornell C, Corradi N (2022) Long reads and Hi-C sequencing illuminate the two-compartment genome of the model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*. New Phytol 233(3):1097–1107. https://doi.org/10.1111/nph.17842

Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y (2012) DbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 40(W1):W445–W451. https://doi.org/10.1093/nar/gks479

Zhang C, Rabiee M, Sayyari E, Mirarab S (2018) ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19(153):16–62. https://doi.org/10.1186/s12859-018-2129-y