

# Early branching arbuscular mycorrhizal fungus *Paraglomus occultum* carries a small and repeat-poor genome compared to relatives in the Glomeromycotina

Mathu Malar C<sup>1</sup>, Yan Wang<sup>2,3</sup>, Jason E. Stajich<sup>4,5</sup>, Vasilis Kokkoris<sup>1</sup>, Matthew Villeneuve-Laroche<sup>1</sup>, Gokalp Yildirir<sup>1</sup> and Nicolas Corradi<sup>1,\*</sup>

## Abstract

The arbuscular mycorrhizal fungi (AMFs) are obligate root symbionts in the subphylum Glomeromycotina that can benefit land plants by increasing their soil nutrient uptake in exchange for photosynthetically fixed carbon sources. To date, annotated genome data from representatives of the AMF orders Glomerales, Diversisporales and Archaeosporales have shown that these organisms have large and highly repeated genomes, and no genes to produce sugars and fatty acids. This led to the hypothesis that the most recent common ancestor (MRCA) of Glomeromycotina was fully dependent on plants for nutrition. Here, we aimed to further test this hypothesis by obtaining annotated genome data from a member of the early diverging order Paraglomerales (*Paraglomus occultum*). Genome analyses showed this species carries a 39.6 Mb genome and considerably fewer genes and repeats compared to most AMF relatives with annotated genomes. Consistent with phylogenies based on ribosomal genes, our phylogenetic analyses suggest *P. occultum* as the earliest diverged branch within Glomeromycotina. Overall, our analyses support the view that the MRCA of Glomeromycotina carried hallmarks of obligate plant biotrophy. The small genome size and content of *P. occultum* could either reflect adaptive reductive processes affecting some early AMF lineages, or indicate that the high gene and repeat family diversity thought to drive AMF adaptability to host and environmental change was not an ancestral feature of these prominent plant symbionts.

## DATA SUMMARY

All genome data is available from the National Center for Biotechnology Information BioProject PRJNA767608 and from Zenodo (<https://doi.org/10.5281/zenodo.5879318>).

## INTRODUCTION

Arbuscular mycorrhizal fungi (AMF) (subphylum Glomeromycotina) [1–3] form symbiotic relationships with over 70% of vascular land plants [4, 5]. The arbuscular mycorrhizal symbiosis (AMS) emerged over 400 million years ago [6], and plays a major role in extant terrestrial ecosystems by allowing plant roots to efficiently acquire poorly soluble soil nutrients (phosphate, nitrogen and other trace elements) in exchange for carbon sources [4]. During AMS, the fungal partner colonizes the plant roots to produce tree-like structures (arbuscules) that promote bidirectional nutrient transfers between partners, including carbohydrates and lipids from plants that AMFs utilize as an energy resource [7]. AMF also interact with the soil microbiome and play a major role in ecosystems and biodiversity by enhancing plant resistance against pathogens, as well as abiotic and biotic stress [8, 9].

Received 06 February 2022; Accepted 10 March 2022; Published 22 April 2022

**Author affiliations:** <sup>1</sup>Department of Biology, University of Ottawa, Ottawa, ON K1N 6N5, Canada; <sup>2</sup>Department of Biological Sciences, University of Toronto Scarborough, Toronto, Canada; <sup>3</sup>Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada; <sup>4</sup>Department of Microbiology and Plant Pathology, University of California–Riverside, Riverside, CA, USA; <sup>5</sup>Institute for Integrative Genome Biology, University of California–Riverside, Riverside, CA, USA.

**\*Correspondence:** Nicolas Corradi, ncorradi@uottawa.ca

**Keywords:** AMF; *Paraglomus occultum*; early diverging.

**Abbreviations:** AMF, arbuscular mycorrhizal fungus; MRCA, most recent common ancestor.

**Data statement:** All supporting data, code and protocols have been provided within the article or through supplementary data files. Supplementary material is available with the online version of this article. Two supplementary figures, eight supplementary files and two supplementary tables are available with the online version of this article.

000810 © 2022 The Authors

 This is an open-access article distributed under the terms of the Creative Commons Attribution NonCommercial License. This article was made open access via a Publish and Read agreement between the Microbiology Society and the corresponding author's institution.

### Impact Statement

Most plants undergo underground symbioses with single celled organisms called arbuscular mycorrhizal fungi (AMF). This symbiosis is prominent in terrestrial ecosystems, where AMFs provide higher access to soil nutrients to plants and represent a food source for multiple organisms. To date, genome data has shown that these fungi carry relatively large and highly repeated genomes, and genome comparisons revealed that AMF have lost the ability to produce sugars and lipids; meaning that these are fully depend on plants for nutrition. In this paper, we aimed to support these findings by sequencing the genome of the early AMF lineage *Paraglomus occultum*. Our analyses confirm that all AMF sequenced to date cannot produce lipids and sugars. However, we found that *P. occultum* carries a very small and gene-poor genome compared to its AMF relatives. This information drastically changes our view that all AMF genomes evolved to be large and repeat rich as a means to improve their adaptability to change.

AMF produce hyphae and spores that harbour thousands of nuclei within one cytoplasm, and studies on model species have revealed that these organisms are either homokaryotic (all coexisting nuclei carry one dominant genotype) or heterokaryotic (nuclei derive from two parental strains) [10, 11]. Genome analyses have also shown that these organisms have lost genes involved in the production of sugars and fatty acids [12–15], and it was recently proposed that these were lost in the most recent common ancestor (MRCA) of this subphylum [16].

Currently, the Glomeromycotina are subdivided into four orders – i.e. Glomerales, Diversisporales, Archaeosporales and Paraglomerales [17]. Among these, species in the Paraglomerales are understudied and currently the only ones without a publicly available annotated genome. The lack of information on the gene repertoires from Paraglomerales primarily results from difficulties in extracting contaminant-free DNA from pot-cultured representatives of this order. To circumvent this, studies have either used poly-A tailed RNA [5] or single nucleus sequencing [18] as a means to remove obvious bacterial and other non-AMF contaminants. While these approaches produced sequence data of sufficient quality to perform phylogenetic analyses, their high fragmentation combined with the absence of gene annotation made these datasets unsuitable for comprehensive comparative genome studies.

Species in the Paraglomerales inhabit most terrestrial environments where other AMF families co-exist. In communities studies, members of this order vary in abundance compared to other AMF taxa, depending on the soil conditions and geographical location [19–22]. There is also evidence that more intensive (disruptive) agricultural practices have direct negative impacts on the abundance of Paraglomerales compared to other species [23, 24], and that species in this order may be less beneficial in stimulating both plant growth and nutrient uptake by roots compared to representatives of the Glomerales and Diversisporales [25]. However, whether any of these traits derive from genetic or genomic traits unique to this AMF order is currently unknown.

To tackle this question, we obtained an annotated draft genome of *Paraglomus occultum* from less than 100 ng genomic DNA isolated from pot-cultured spores using a *k*-mer based [26] metagenomics approach. Our phylogenetic and comparative analyses revealed that this species carries a small and streamlined genome size and content compared to Glomeromycotina relatives, and suggest that Paraglomerales could represent the earliest phylogenetic node of this group of prominent symbionts.

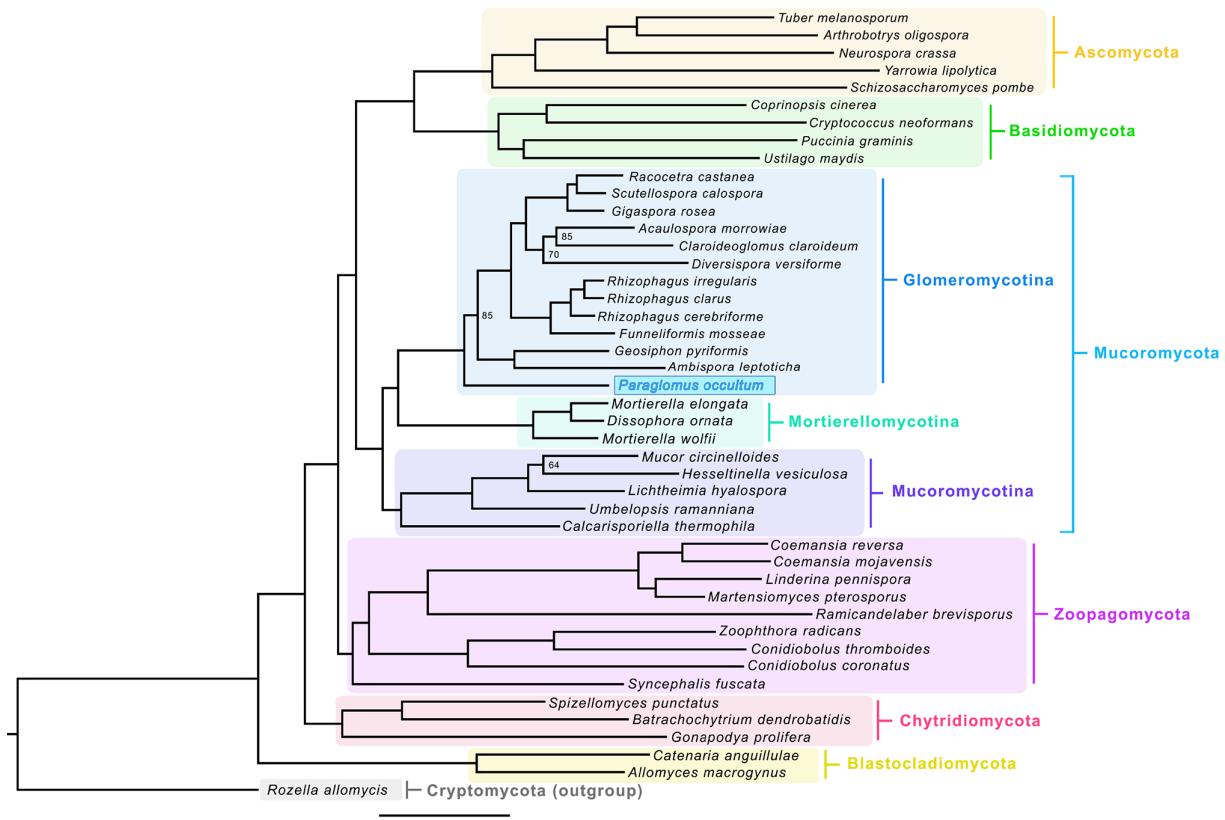
## METHODS

### Cultivation of *P. occultum* samples

Quiescent spores from *P. occultum* (DAOM 240472) were harvested from pot-cultures available at the Glomeromycotina *in vivo* pot-culture collection at Agriculture and Agri-Food Canada (AAFC) (Ottawa, ON, Canada). This fungal collection is regularly checked for contamination using both morphological and molecular techniques (i.e. small subunit of the rRNA gene, SSU; H+ATPases). Using an inverted microscope, only quiescent spores showing viable morphology and lipid contents were used for downstream DNA extraction. Spores were collected with a sieve cascade of descending size ranging from 500 to 63 mm, and the lowest fraction was then placed in a 50 ml Falcon tube with the addition of 50% (w/w) sucrose solution. Following centrifugation at 5000 g for 5 min, single spores were isolated from the supernatant, and general species identification was confirmed using morphological descriptions as detailed by Beaudet *et al.* [5]. Spores were surface sterilized using successive baths of chloramine-T, 2% Tween 20 for 2 min, and sterile distilled water for 1 min, then incubated overnight in a solution of 0.2 mg streptomycin ml<sup>-1</sup> and 0.1 mg gentamicin ml<sup>-1</sup>, and washed in a final bath of sterile distilled water. Sterilized spores were stored at 4 °C.

### DNA extraction and genome sequencing and assembly

DNA was extracted from 250 individual spores of *P. occultum* using the Qiagen DNA extraction kit, following the manufacturer's procedures. This resulted in the extraction of 2 ng DNA. The DNA was sent to Fasteris to generate Nextera Illumina paired-end libraries, which were sequenced using the HiSeq 2500 platform. This procedure generated over 12 billion base pairs of sequence



**Fig. 1.** Phylogenetic tree showing the placement of *P. occultum* in the Glomeromycotina clade. The tree was resolved using IQ-TREE on a concatenated alignment of 434 protein-encoding genes. Numbers indicate nodes with less than 95% bootstrap support. Branches are coloured according to their phylum. The scale bar indicates number of substitutions per site.

data with 93% of base pairs exceeding a quality score greater than or equal to 30 (Q30). Poor quality and adapter sequences were trimmed using Trimmomatic with the following parameters of ILLUMINACLIP: 2:30:10 SLIDINGWINDOW: 5:20 LEADING:5 TRAILING:5 MINLEN:50.

The resulting reads were assembled using metaSPAdes v3.11.1 [27], and assembled contigs were binned based on tetra-nucleotide signatures using CONCOCT [28], as described previously [16]. Binned clusters were annotated using the BLAST v 2.6.0+ [29, 30] nr database with *E* value cut-off  $1 \times 10^{-5}$ , and clusters matching to bacteria were removed. The retained AMF contigs were used as a reference database to filter the total paired-end reads with BWA [31] using default parameters, where reads mapping to the filtered contigs were extracted to construct a cleaned paired-end sequence library. These cleaned libraries were *de novo* assembled with SPAdes v3.11.1 [32] with parameters of -k 21,33,55,77 --careful. An additional round of TBLASTN searches of the SPAdes assembled contigs against the National Center for Biotechnology Information (NCBI) nr database was performed to remove detected further bacterial contamination. K-mer ( $k=21$ ) was used on cleaned and filtered reads to estimate the genome size of *P. occultum* using jellyfish 1.1.12 [33].

## Genome annotation

The assembled genome was soft masked before gene prediction using TANTAN [34], and genome annotation was performed using Funannotate v1.7.4 (<https://funannotate.readthedocs.io/>) [<https://doi.org/10.5281/zenodo.3679386>]. Gene annotation was carried out using published *Paraglomus* RNA transcripts and RNA-seq data [5]. Gene functions were identified using Diamond BLASTX [30], Pfam domain analysis was performed using Pfamscan [35, 36]. Carbohydrate-active enzymes (CAZymes) were identified using the dbCAN2 server [37, 38] and secretory proteins using published pipelines [14, 37]. Metabolic pathway analysis was performed using the KAAS KEGG server [39], and eggNOG 5.0 [40] was used to classify the orthologue gene functions of Mortierellomycotina and Glomeromycotina genomes used in this study. Transposable elements were predicted using TransposonPSI [41]. The completeness of the genome assembly was assessed with BUSCO version 4.0 [42] on protein datasets with default parameters using the fungal gene dataset [fungi\_odb10].

**Table 1.** Genome assembly statistics of *P. occultum* with other Mucoromycota genomes used in this study

Genome	Assembly size (Mb)	No. of scaffolds	Scaffold N50	Largest scaffold (kb)	Repeat %	busco completeness % using V4	G+C (mol%)	Gene density per 10 kb
<i>Paraglomus occultum</i>	44	4722	160	104.458	12.78	92.2	36.77	2.276
<i>Geosiphon pyriformis</i>	129	795	703	2733.91	64.35	93.2	29.25	1.893
<i>Gigaspora rosea</i> V1.0	597.95	7526	734	1204.75	63.44	96.3	28.81	0.550
<i>Rhizophagus cerebriforme</i> DAOM227022 V1.0	136.89	2592	266	709.02	24.77	97.2	26.55	1.574
<i>Rhizophagus irregularis</i> DAOM 197198V2.0	136.80	1123	129	1375.86	26.38	96.6	27.53	1.913
<i>Diversispora versiformis</i> strain IT104	147	731	434	2010.39	43.6	95.5	25.1	1.810
<i>Rhizopus microsporus</i> ATCC 11559 V1	25.97	131	8	2782.17	4.68	95.7	37.48	4.610
<i>Mucor circinelloides</i> CBS 277.49 V2.0	36.59	26	4	6050.25	20.38	95.3	42.17	3.194
<i>Phycomyces blakesleeanus</i> NRRL1555 V2.0	53.94	80	11	4452.46	9.74	95.7	35.78	3.062
<i>Mortierella elongata</i> AG-77	49.86	473	31	1526.29	4.63	99	48.05	3.002

## Protein orthology analyses

For comparative analyses, published genomes from Glomeromycotina, Mortierellomycotina and Mucoromycotina (10 species in total) were downloaded from the Joint Genome Institute (JGI) portal MycoCosm database [43, 44], and the predicted proteins were clustered with FASTORTHO using 50% identify and coverage [14]. From the clustered orthogroup, orthologues were classified into conserved genes (proteins shared by all genomes), dispensable genes (proteins that are present in at least two genomes), and species-specific genes (taxonomically restricted proteins). For each category, we also identified duplicated genes.

Single copy core genes were aligned using MAFFT [45] with default parameters and the resulting alignments were trimmed to remove positions with gaps using trimAl v1.4. rev22 [46]. A maximum-likelihood tree was reconstructed using IQ-TREE v1.6.12 [47] with 10 000 replicates of SH-aLRT. Trees were visualized and annotated with FigTree v1.4.4 (<https://beast.community/figtree>).

## Phylogenetic analysis and alternative topology test

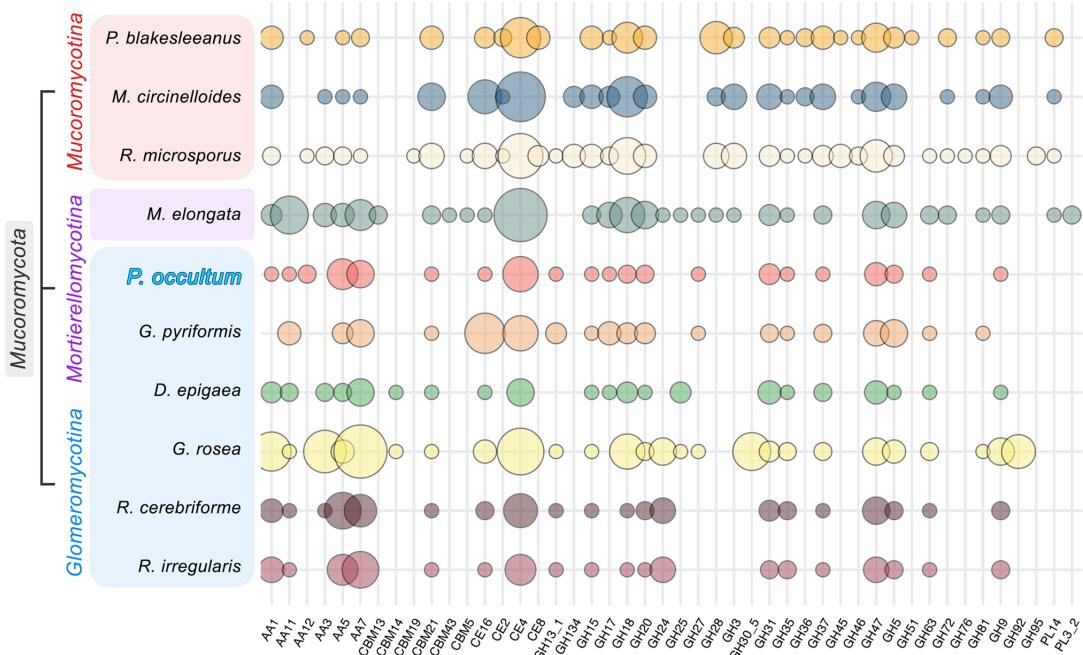
Phylogenomic analysis was carried out using the 'fungi\_odb10' dataset included in the busco v4.0 package. Profile hidden-Markov-models of these markers were used to identify homologues in *P. occultum* and an additional 44 fungal genomes using HMMER3 (v3.1b2) employed in the PHYling pipeline (<https://doi.org/10.5281/zenodo.1257002>). A total of 603 conserved and single-copy proteins were identified, aligned and concatenated for the following phylogenetic analyses. The phylogenetic tree was reconstructed using the maximum-likelihood approach with the best-fit substitution model option for individual partitions defined by each gene marker implemented in IQ-TREE v.1.6. The final analysis was performed on 195323 sites with 170865 distinct patterns contributing to the tree structures.

To test the likelihood of alternative phylogenetic placements of *P. occultum*, we conducted constraint tree searches using partially forced tree topology files (Alt\_T1, Alt\_T2 and Alt\_T3 as illustrated in Fig. S1, available with the online version of this article) with the '-g' option implemented in the IQ-TREE package. Tree files of the best topology (as shown in Fig. 1) and the alternative ones (Fig. S1) were included in the computational analysis for log-likelihoods using Kishino–Hasegawa test, Shimodaira–Hasegawa test, expected likelihood weight, and approximately unbiased test via the '-zb' and '-au' parameters included in the IQ-TREE package [48–52]. Reliable results were achieved using 10000 resampling estimated log-likelihood (RELL) replicates.

## RESULTS

### Genomic characteristics of *P. occultum*

DNA isolated from *P. occultum* spores was sequenced using two Illumina libraries with read lengths of 125 bp, producing a total of 54 million paired-end reads. The genome was assembled using the *k*-mer binning based approach recently used to acquire the



**Fig. 2.** Plot showing the count, presence and absence of plant cell wall degrading enzyme (PCWDE) categories. The x-axis shows the categories of PCWDE-encoding enzymes and the y-axis shows the list of Mucoromycota species used in this genomics study. Species highlighted in orange represent Mucoromycotina, purple highlighted species are from Mortierellomycotina, and the blue colour represents the species from the Glomeromycotina clade.

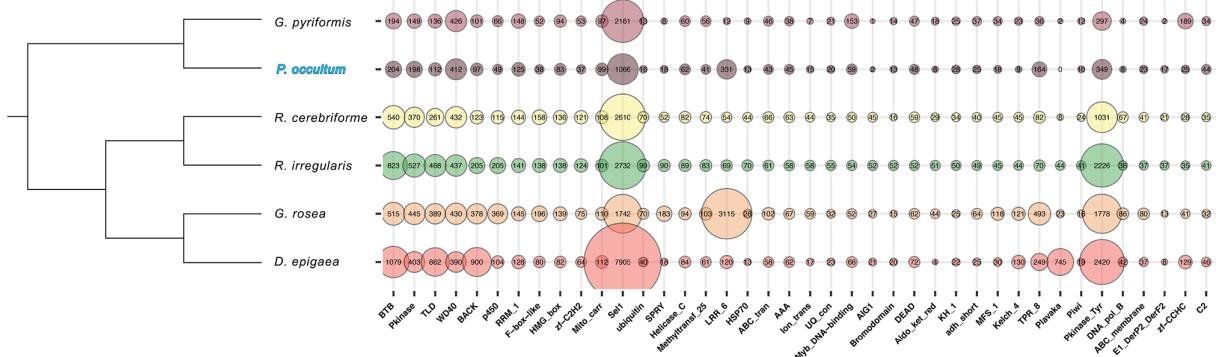
genome of *Geosiphon pyriformis* [16] and *Diversispora epigaea* [13]. This analysis generated 4722 scaffolds for *P. occultum* with a total length of 44 Mb, which compares to sizes of 140 Mb and above for publicly available annotated genomes. In total, 10145 genes were predicted in the genome, which compares to a mean of 24000 genes for other AMF taxa, and the *P. occultum* gene density and G+C content is significantly higher than other AMFs (Table 1).

Overall, the *P. occultum* genome size and gene counts are approximately a third of those found in other AMF. While this small size may suggest that a large portion of the *P. occultum* genome was missed/lost during the assembly process, BUSCO completeness analyses were virtually identical to those found in other sequenced non-model AMF genomes (92.2%; Table 1) [13–16]. The reduced genome size and repeat content we observed is also validated by assembly sizes obtained by others on the same species and *Paraglomus brasiliannum* using independent approaches – i.e. single nucleus sequencing (49 Mb in the work by Montoliu-Nerin et al. [18] vs 44 Mb for this study).

Our secretome analysis revealed the presence of 209 putative secreted proteins (Table S1) (Supplementary Material 1) and 120 CAZymes in this species (File S1). The genome also contains 52 plant cell wall degrading enzymes, on a par with numbers found in other AMF species (Fig. 2, File S2). Like all sequenced AMFs, *P. occultum* lacks the 'Missing Glomeromycotina Core Genes' (MGCGs), including those involved in the production of fatty acids and sugars (File S3). Our data also indicates this species is likely haploid and homokaryotic – i.e. mapping reads onto the *P. occultum* genome does not reveal evidence of two distinct genotypes [10, 53, 54]. The genome also carries a complete set of meiosis-specific genes (Table S2), further supporting the presence of cryptic (para)sexual reproduction across the subphylum [55]. Analyses of metabolic pathways show that *P. occultum* carries genes involved in nitrogen metabolism and associated transporters in numbers similar to those found in AMF relatives (File S4).

### Phylogenetic placement of *P. occultum*

A set of 603 conserved single-copy genes that includes those from *P. occultum* was used to reconstruct a phylogenetic tree of the fungal kingdom (Fig. 1). These analyses fully support the monophyly of Glomeromycotina and its close relationship with Mortierellomycotina within the phylum Mucoromycota [3, 56]. Remarkably, in support of rRNA-based phylogenies [57, 58], but in minor disagreement with recent protein-code-based work [5, 16, 18], our analyses place *P. occultum* as the earliest lineage within the Glomeromycotina phylogeny [5, 16, 18]. Furthermore, our analyses place *Claroideoglomus claroideum* (*Claroideoglomus* is a genus that is notoriously difficult to assess phylogenetically) within Diversisporales, while other studies placed this genus either as a sister clade to Diversisporales/Glomerales [16, 18] or as a member of Glomerales [57, 58].



**Fig. 3.** Phylogenetic tree with bubble plot showing the comparison of all the top 40 Pfam domains of AMF genomes used in this study. The x-axis represents Pfam functional domains and the y-axis represents the species of Glomeromycotina genomes.

The placement of *P. occultum* in our analyses has full statistical support. Alternative topologies that forced this species as a sister lineage to the *Geosiphon-Ambispora* clade (Alt-T1; Fig. S1) or as being associated with Glomerales or Diversisporales (Alt-T2, Alt-T3; Fig. S1) were either rejected (Alt\_T2 and Alt\_T3) or less favoured (Alt\_T1) using statistical tests implemented in IQ-TREE, including the KH, SH, ELW and AU tests (File S5).

#### Shared and unique gene and genome characteristics of *P. occultum*

Glomeromycotina genomes are expected to be large relative to most fungi, and carry large gene family expansions [11, 12, 14, 59]. *P. occultum* carries gene expansions of the tetratricopeptide repeat Sel1-, the homodimerization BTB (broad-complex, tramtrack and bric-a-brac) and WD-40 domain-containing protein domains (Fig. 3, File S6). However, in this species, most of these domains are less abundant compared to other AMF taxa with annotated genomes. For example, *P. occultum* carries 1066 Sel1 and 349 Pkinase Tyr motifs, compared to a mean of 3400 and 1500 for other Glomeromycotina with sequenced genomes (Fig. 3, File S6). Despite its relatively smaller number of protein domains, the *P. occultum* genome shows enrichment in My-DNA binding domains and zinc-finger motifs.

*P. occultum* also contains a fraction of transposable elements found in Glomeromycotina relatives – e.g. 882 (or 0.32 per Mb) transposable element families compared to a mean of 16400 (or 10.17 per Mb) for other Glomeromycotina. These numbers are similar (or higher) to those found in members of Mucoromycotina and Mortierellomycotina with similar genome sizes (Fig. S2, File S7). By comparing our data with genome size and repeat estimates recently obtained from other early AMF taxa, including another member of Paraglomerales (*P. brasiliense*) [18], we found that genome size in AMF significantly correlates with repeat content in the Glomeromycotina (Fig. 4).

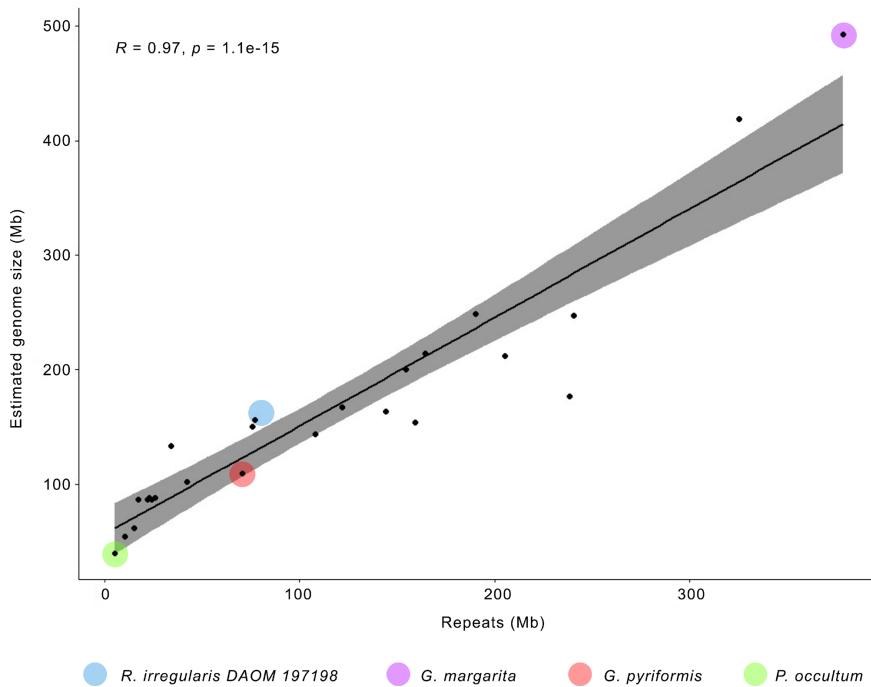
To identify core and unique Glomeromycotina genes, we clustered the predicted protein families of six Glomeromycotina and four Mucoromycotina genomes to detect orthologous groups. Consistent with recent studies [14], orthologues can be classified into core genes (shared among all Mucoromycota), dispensable genes (shared by at least two species of Mucoromycotina or Glomeromycotina species), and species-specific genes (Fig. 5). Due to its small genome size, *P. occultum* contains more than double the percentage of core genes (23%, compared to ~10% for other Glomeromycotina) and, on a par with values found in Mortierellomycotina, significantly less species-specific and dispensable genes (39% compared to ~50%) (Fig. 5). *P. occultum* also has 200 gene orthologue losses involved in a variety of putative functions (File S8).

## DISCUSSION

#### Paraglomerales may represent the earliest AMF phylogenetic node

To date, AMF taxonomy is primarily generated by linking a small number of morphological traits (e.g. spore structure and content) with phylogenetic analyses of the small subunit of the rRNA operon [17, 60]. These rRNA-based analyses supported the existence of four AMF orders – Diversisporales, Glomerales, Archaeosporales and Paraglomerales – and in these the Paraglomerales represent the earliest lineage [61]. In recent years, the advent of new sequencing methods generated protein-encoding sequence data from poorly studied lineages, and its use in phylogenomics produced a different phylogenetic scenario in which the Paraglomerales group is in a clade that contains members of the Archaeosporales (*Ambispora* and *Geosiphon*) [5, 18, 23, 24].

In this study, an extended orthologue gene set extracted from the *P. occultum* genome produced phylogenies surprisingly consistent with those obtained using rRNA sequences. However, we were also unable to reject an alternative placement of *P. occultum* with members of the Archaeosporales. As such, both current placements of Paraglomerales within the Glomeromycotina phylogeny



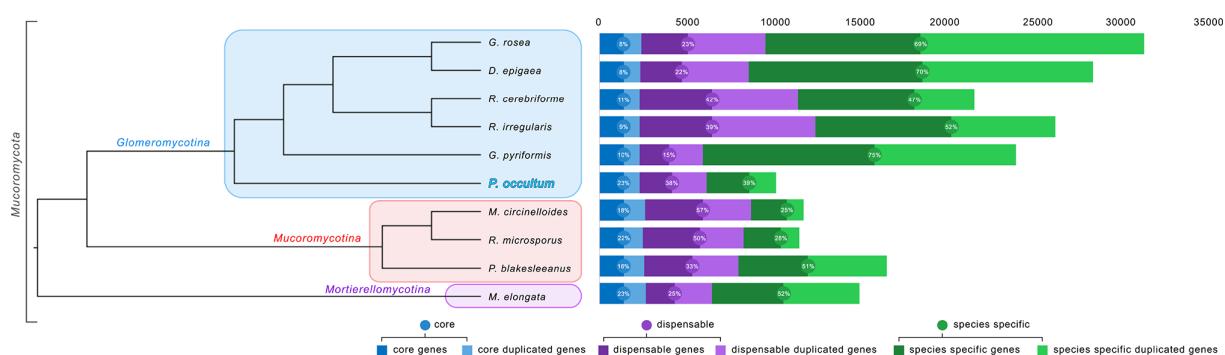
**Fig. 4.** Relationship between repeat content and estimated genome size. All AMF species from Merce et al. (2021) [18] are used here along the assembled genome of *P. occultum*. The blue circle represents *Rhizophagus irregularis* and the purple circle represents *Gigaspora margarita* from Merce et al. (2021) [18]; the red circle represents *Geosiphon pyriformis* from Malar C et al. (2021) [16], and the green circle represent the *P. occultum* genome from this study.

– i.e. earliest node, or with members of Archaeospora – are equally supported by current phylogenetic analyses, and the acquisition and analysis of annotated genome data from early diverging AMF species will hopefully clarify those relationships.

#### A streamlined genome in the MRCA of all extant Glomeromycotina or lineage-specific gene losses?

The present work showed that the *P. occultum* genome has a size, and gene and repeat counts more similar to members of the Mortierellomycotina – i.e. the sister clade of Glomeromycotina – than to most AMF species. These findings are supported by BUSCO values indicating that the *P. occultum* genome space is largely sampled, and by reports of virtually identical results obtained on two members of the Paraglomerales (*P. occultum* and *P. brasiliense*) using independent approaches [18].

Those genome features, combined with the early placement of *P. occultum* in the phylogeny, suggest two contrasting scenarios for Glomeromycotina genome evolution. On one hand, the streamlined features of the *P. occultum* genome could indicate



**Fig. 5.** Organismal phylogeny with bar graphs representing conserved proteins shared among the ten species: core proteins (dark blue colour), duplicated core proteins (light blue colour), dispensable proteins (dark purple colour), duplicated dispensable proteins (light purple colour), species-specific proteins (dark green colour), and duplicated species-specific proteins (light green colour).

that the MRCA of Glomeromycotina was an obligate plant biotroph but also gene and repeat poor. If correct, then gene and repeat family expansions commonly attributed to these prominent plant symbionts must have emerged later in the more derived AMF lineages, possibly to improve their adaptability to environmental and host change [11, 62, 63]. On the other hand, it is possible that the Paraglomerales have undergone drastic and species-specific genome reductive processes and the observation of over 200 *P. occultum* specific gene losses support this view. Either way, genome evidence from additional members of this order is needed to determine which scenario is more likely.

From an ecological perspective, it is intriguing to speculate that the gene family reduction we observed is linked to the reduced mycorrhizal capabilities recently reported for this order in comparison to AMF relatives [23, 24]. Alternatively, the smaller genome might be driven by selection for fast dividing life stages, adaptation to nitrogen limitation, or there may be due to unknown conditions where Paraglomerales are relatively more abundant (e.g. grow better than other Glomeromycota under high metal toxicity, or very high or very low pH, or salinity, etc.). Current analyses did not reveal evidence of gene expansions related to nitrate or metal metabolism and transport in *P. occultum* that could support this hypothesis, however, and data on the nuclear dynamics/division of early AMF lineages are currently non-existent [64].

## Conclusions

The present findings revealed that not all Glomeromycotina carry large and repetitive genomes. In fact, we show that early diverged members of this group can carry streamlined genomes with sizes and contents similar to those of other Mucoromycota subphyla. These reductive features can be explained by two contrasting hypotheses about the evolution of Glomeromycotina – i.e. late genome expansions vs lineage-specific adaptive reduction – and future studies should now focus on adding new annotated genome data from strains that belong to early lineages (including Paraglomerales) to address these conflicting views. Lastly, this work demonstrates that a complete, contaminant-free view of the genome content can be obtained from pot-cultured AMFs using a few nanograms of DNA, and without the need for expensive DNA/RNA extraction kits tailored to small amounts of starting material [5], or instruments dedicated to the isolation and sequencing of individual nuclei [54, 65].

### Funding information

Our research is funded by the Discovery programme of the Natural Sciences and Engineering Research Council (RGPIN2020-05643), a Discovery Accelerator Supplements programme (RGPAS-2020-00033). N.C. is a University of Ottawa Research Chair in Microbial Genomics. J.E.S. is a CIFAR Fellow in the programme Fungal Kingdom: Threats and Opportunities. J.E.S. was supported by US National Science Foundation grants DEB-1441715 and DEB-1557110.

### Conflicts of interest

The authors declare that there are no conflicts of interest.

### References

1. Schüßler A, Schwarzott D, Walker C. A new fungal phylum, the glomeromycota: phylogeny and evolution. *Mycol Res* 2001;105:1413–1421.
2. Schüßler A, Walker C. Evolution of the 'plant-symbiotic' fungal phylum, Glomeromycota. In: *The Mycota – Evolution of Fungi and Fungal-Like Organisms*. Heidelberg: Springer, 2011. pp. 163–185.
3. Spatafora JW, Chang Y, Benny GL, Lazarus K, Smith ME, et al. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* 2016;108:1028–1046.
4. Smith SE, Read DJ. *Mycorrhizal Symbiosis*. 3rd edn. San Diego and London: Academic Press; 2008.
5. Beaudet D, Chen ECH, Mathieu S, Yildirir G, Ndikumana S, et al. Ultra-low input transcriptomics reveal the spore functional content and phylogenetic affiliations of poorly studied arbuscular mycorrhizal fungi. *DNA Res* 2018;25:217–227.
6. Remy W, Taylor TN, Hass H, Kerp H. Four hundred-million-year-old vesicular arbuscular mycorrhizae. *Proc Natl Acad Sci USA* 1994;91:11841–11843.
7. Parniske M. Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nat Rev Microbiol* 2008;6:763–775.
8. Newsham KK, Fitter AH, Watkinson AR. Arbuscular mycorrhiza protect an annual grass from root pathogenic fungi in the field. *J Ecol* 1995;83:991–1000.
9. Boomsma CR, Vyn TJ. Maize drought tolerance: potential improvements through arbuscular mycorrhizal symbiosis? *Field Crops Res* 2008;108:14–31.
10. Ropars J, Toro KS, Noel J, Pelin A, Charron P, et al. Evidence for the sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. *Nat Microbiol* 2016;1:16033.
11. Chen ECH, Morin E, Beaudet D, Noel J, Yildirir G, et al. High intraspecific genome diversity in the model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*. *New Phytol* 2018;220:1161–1171.
12. Tisserant E, Malbreil M, Kuo A, Kohler A, Symeonidi A, et al. Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proc Natl Acad Sci USA* 2013;110:20117–20122.
13. Sun X, Chen W, Ivanov S, MacLean AM, Wight H, et al. Genome and evolution of the arbuscular mycorrhizal fungus *Diversispora epigaea* (formerly *Glomus versiforme*) and its bacterial endosymbionts. *New Phytol* 2019;221:1556–1573.
14. Morin E, Miyauchi S, San Clemente H, Chen ECH, Pelin A, et al. Comparative genomics of *Rhizophagus irregularis*, *R. cerebriforme*, *R. diaphanus* and *Gigaspora rosea* highlights specific genetic features in Glomeromycotina. *New Phytol* 2019;222:1584–1598.
15. Kobayashi Y, Maeda T, Yamaguchi K, Kameoka H, Tanaka S, et al. The genome of *Rhizophagus clarus* HR1 reveals a common genetic basis for auxotrophy among arbuscular mycorrhizal fungi. *BMC Genomics* 2018;19:465.
16. Malar C M, Krüger M, Krüger C, Wang Y, Stajich JE, et al. The genome of *Geosiphon pyriformis* reveals ancestral traits linked to the emergence of the arbuscular mycorrhizal symbiosis. *Curr Biol* 2021;31:1570–1577.

17. Redecker D, Schüssler A, Stockinger H, Stürmer SL, Morton JB, et al. An evidence-based consensus for the classification of arbuscular mycorrhizal fungi (Glomeromycota). *Mycorrhiza* 2013;23:515–531.
18. Montoliu-Nerin M, Sánchez-García M, Bergin C, Kutschera VE, Johannesson H, et al. In-depth phylogenomic analysis of arbuscular mycorrhizal fungi based on a comprehensive set of *de novo* genome assemblies. *Front Fungal Biol* 2021;2:53.
19. Yang W, Zhang M, Song F, Liu S, Li X, et al. Comparative analysis of arbuscular mycorrhizal fungal communities between farmland and woodland in the black soil region of Northeast China. *Agric* 2021;11:866.
20. Casazza G, Lumini E, Ercole E, Dovana F, Guerrina M, et al. The abundance and diversity of arbuscular mycorrhizal fungi are linked to the soil chemistry of screes and to slope in the Alpic paleo-endemic *Berardia subacaulis*. *PLoS One* 2017;12:e0171866.
21. Barceló M, van Bodegom PM, Tedersoo L, den Haan N, Veen GFC, et al. The abundance of arbuscular mycorrhiza in soils is linked to the total length of roots colonized at ecosystem level. *PLoS One* 2020;15:e0237256.
22. Zhu X, Yang W, Song F, Li X. Diversity and composition of arbuscular mycorrhizal fungal communities in the cropland black soils of China. *Glob Ecol Conserv* 2020;22:e00964.
23. Gosling P, Proctor M, Jones J, Bending GD. Distribution and diversity of *Paraglomus* spp. in tilled agricultural soils. *Mycorrhiza* 2014;24:1–11. 2013 May 30.
24. Stefani F, Bencherif K, Sabourin S, Hadj-Sahraoui AL, Banchini C, et al. Taxonomic assignment of arbuscular mycorrhizal fungi in an 18S metagenomic dataset: a case study with saltcedar (*Tamarix aphylla*). *Mycorrhiza* 2020;30:243–255.
25. Säle V, Palenzuela J, Azcón-Aguilar C, Sánchez-Castro I, Alves da Silva G, et al. Ancient lineages of arbuscular mycorrhizal fungi provide little plant benefit. *Mycorrhiza* 2021;31:559–576.
26. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, et al. Binning metagenomic contigs by coverage and composition. *Nat Methods* 2014;11:1144–1146.
27. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 2017;27:824–834.
28. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, et al. Binning metagenomic contigs by coverage and composition. *Nat Methods* 2014;11:1144–1146.
29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
30. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2015;12:59–60.
31. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–1760.
32. Bankevich A, Nurk S, Antipov D, Gurevich AA, Divorkin M, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
33. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 2011;27:764–770.
34. Frith MC. A new repeat-masking method enables specific detection of homologous sequences. *Nucleic Acids Res* 2011;39:e23.
35. Li W, Cowley A, Uludag M, Gur T, McWilliam H, et al. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res* 2015;43:W580–W584.
36. Hancock JM, Zvelebil MJ, Hancock JM, Bishop MJ. HMMer. In: *Dictionary of Bioinformatics and Computational Biology*. Hoboken, NJ: Wiley, 2004.
37. Pellegrin C, Morin E, Martin FM, Veneault-Fourrey C. Comparative analysis of secretomes from ectomycorrhizal fungi with an emphasis on small-secreted proteins. *Front Microbiol* 2015;6:1278.
38. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 2014;42:D490–D495.
39. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 2007;35:W182–W185.
40. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forsslund SK, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 2019;47:D309–D314.
41. Hass B. *TransposonPSI: an Application of PSI-Blast to Mine (Retro-) Transposon ORF Homologies*. Cambridge, MA: Broad Institute; 2010.
42. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015;31:3210–3212.
43. Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, et al. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res* 2014;42:D699–D704.
44. Grigoriev IV, Nordberg H, Shabalov I, Aerts A, Cantor M, et al. The genome portal of the Department of Energy Joint Genome Institute. *Nucleic Acids Res* 2012;40:D26–D32.
45. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.
46. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;25:1972–1973.
47. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32:268–274.
48. Kishino H, Hasegawa M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J Mol Evol* 1989;29:170–179.
49. Kishino H, Miyata T, Hasegawa M. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J Mol Evol* 1990;31:151–160.
50. Shimodaira H, Hasegawa M. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* 1999;16:1114–1116.
51. Shimodaira H. An approximately unbiased test of phylogenetic tree selection. *Syst Biol* 2002;51:492–508.
52. Strimmer K, Rambaut A. Inferring confidence sets of possibly misspecified gene trees. *Proc Biol Sci* 2002;269:137–142.
53. Chen ECH, Mathieu S, Hoffrichter A, Ropars J, Dreissig S, et al. More filtering on SNP calling does not remove evidence of inter-nucleus recombination in dikaryotic arbuscular mycorrhizal fungi. *Front Plant Sci* 2020;11:912.
54. Chen E, Mathieu S, Hoffrichter A, Ropars J, Dreissig S, et al. Single nucleus sequencing reveals evidence of inter-nucleus recombination in arbuscular mycorrhizal fungi. *Elife* 2019;7:912.
55. Yıldırır G, Malar C M, Kokkoris V, Corradi N. Parasexual and sexual reproduction in arbuscular mycorrhizal fungi: room for both. *Trends Microbiol* 2020;28:517–519.
56. Spatafora JW, Aime MC, Grigoriev IV, Martin F, Stajich JE, et al. The fungal tree of life: from molecular systematics to genome-scale phylogenies. *Microbial Sprectrum* 2017;5:5.5.03.
57. Krüger M, Krüger C, Walker C, Stockinger H, Schüßler A. Phylogenetic reference data for systematics and phylotaxonomy of arbuscular mycorrhizal fungi from phylum to species level. *New Phytol* 2012;193:970–984.
58. Krüger M, Stockinger H, Krüger C, Schüßler A. DNA-based species level detection of Glomeromycota: one PCR primer set for all arbuscular mycorrhizal fungi. *New Phytol* 2009;183:212–223.
59. Mathieu S, Cusant L, Roux C, Corradi N. Arbuscular mycorrhizal fungi: intraspecific diversity and pangenomes. *New Phytol* 2018;220:1129–1134.
60. Morton JB, Msiska Z. Phylogenies from genetic and morphological characters do not support a revision of *Gigasporaceae*

(Glomeromycota) into four families and five genera. *Mycorrhiza* 2010;20:483–496.

61. Krüger M, Krüger C, Walker C, Stockinger H, Schüßler A. Phylogenetic reference data for systematics and phylotaxonomy of arbuscular mycorrhizal fungi from phylum to species level. *New Phytol* 2012;193:970–984.
62. Yildirir G, Sperschneider J, Malar C M, Chen ECH, Iwasaki W, et al. Long reads and Hi-C sequencing illuminate the two-compartment genome of the model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*. *New Phytol* 2022;233:1097–1107.
63. Reinhardt D, Roux C, Corradi N, Di Pietro A. Lineage-specific genes and cryptic sex: parallels and differences between arbuscular mycorrhizal fungi and fungal pathogens. *Trends Plant Sci* 2021;26:111–123.
64. Kokkoris V, Stefani F, Dalpé Y, Dettman J, Corradi N. Nuclear dynamics in the arbuscular mycorrhizal fungi. *Trends Plant Sci* 2020;25:765–778.
65. Montoliu-Nerin M, Sánchez-García M, Bergin C, Grabherr M, Ellis B, et al. Building *de novo* reference genome assemblies of complex eukaryotic microorganisms from single nuclei. *Sci Rep* 2020;10:1303.

**Five reasons to publish your next article with a Microbiology Society journal**

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

**Find out more and submit your article at [microbiologyresearch.org](https://microbiologyresearch.org).**