



In-depth Phylogenomic Analysis of Arbuscular Mycorrhizal Fungi Based on a Comprehensive Set of *de novo* Genome Assemblies

Merce Montoliu-Nerin^{1*†}, Marisol Sánchez-García^{1,2†}, Claudia Bergin³, Verena Esther Kutschera⁴, Hanna Johannesson⁵, James D. Bever⁶ and Anna Rosling^{1*}

¹ Department of Ecology and Genetics, Evolutionary Biology, Uppsala University, Uppsala, Sweden, ² Department of Forest Mycology and Plant Pathology, Uppsala Biocentre, Swedish University of Agricultural Sciences, Uppsala, Sweden, ³ Microbial Single Cell Genomics Facility, Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden, ⁴ Department of Biochemistry and Biophysics, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Stockholm University, Solna, Sweden, ⁵ Department of Organismal Biology, Systematic Biology, Uppsala University, Uppsala, Sweden, ⁶ Department of Ecology and Evolutionary Biology, and Kansas Biological Survey, University of Kansas, Lawrence, KS, United States

OPEN ACCESS

Edited by:

Jun-Yi Leu,
Academia Sinica, Taiwan

Reviewed by:

Huei-Mien Ke,
Academia Sinica, Taiwan
Ying Chang,
Oregon State University, United States

*Correspondence:

Merce Montoliu-Nerin
merce.montoliu@gmail.com
Anna Rosling
anna.rosling@ebc.uu.se

[†]These authors share first authorship

Specialty section:

This article was submitted to
Fungal Genomics and Evolution,
a section of the journal
Frontiers in Fungal Biology

Received: 28 May 2021

Accepted: 06 September 2021

Published: 29 September 2021

Citation:

Montoliu-Nerin M, Sánchez-García M,
Bergin C, Kutschera VE,
Johannesson H, Bever JD and
Rosling A (2021) In-depth
Phylogenomic Analysis of Arbuscular
Mycorrhizal Fungi Based on a
Comprehensive Set of *de novo*
Genome Assemblies.
Front. Fungal Biol. 2:716385.
doi: 10.3389/fpub.2021.716385

Morphological characters and nuclear ribosomal DNA (rDNA) phylogenies have so far been the basis of the current classifications of arbuscular mycorrhizal (AM) fungi. Improved understanding of the evolutionary history of AM fungi requires extensive ortholog sampling and analyses of genome and transcriptome data from a wide range of taxa. To circumvent the need for axenic culturing of AM fungi we gathered and combined genomic data from single nuclei to generate *de novo* genome assemblies covering seven families of AM fungi. We successfully sequenced the genomes of 15 AM fungal species for which genome data was not previously available. Comparative analysis of the previously published *Rhizophagus irregularis* DAOM197198 assembly confirm that our novel workflow generates genome assemblies suitable for phylogenomic analysis. Predicted genes of our assemblies, together with published protein sequences of AM fungi and their sister clades, were used for phylogenomic analyses. We evaluated the phylogenetic placement of Glomeromycota in relation to its sister phyla (Mucoromycota and Mortierellomycota), and found no support to reject a polytomy. Finally, we explored the phylogenetic relationships within Glomeromycota. Our results support family level classification from previous phylogenetic studies, and the polyphyly of the order Glomerales with Claroideoglomeraceae as the sister group to Glomeraceae and Diversisporales.

Keywords: genomics, phylogenetic, single nuclei sequencing, topology, Glomeromycota

INTRODUCTION

Arbuscular mycorrhizal (AM) fungi are an ecologically important group of fungi that form ubiquitous associations with plants, establishing symbiosis with up to 80% of land plant species (Parniske, 2008; Smith and Read, 2010). Arbuscular mycorrhizal fungi play foundational roles in terrestrial productivity, and there is accumulating evidence that AM fungal taxa are functionally

distinct and that their community composition have functional consequences for terrestrial ecosystems (Hoeksema et al., 2018; Koziol et al., 2018). Therefore, progress in understanding the ecologically distinct roles of AM fungi depends upon accurate phylogenetic inference at all taxonomic levels.

Available literature identify that all AM fungi form a monophyletic lineage within the fungal kingdom. This lineage is taxonomically classified either as a phylum, Glomeromycota (Schüßler et al., 2001; Hibbett et al., 2007; Schüßler and Walker, 2010; Tedersoo et al., 2018), or as the sub-phylum Glomeromycotina, which together with Mortierellomycotina and Mucoromycotina, make up the phylum Mucoromycota (Spatafora et al., 2016; James et al., 2020; Li et al., 2021). The current consensus classification of AM-fungal species into genera and families was established by Redecker et al. in 2013, when systematists with long experience in the biology and taxonomy of AM fungi joined forces to integrate morphological and molecular phylogenetic information to generate a meaningful classification that reflects evolutionary relationships (Redecker et al., 2013). Molecular data at the time was primarily based on partial sequences of nuclear ribosomal DNA (rDNA). Around 300 species of AM fungi are currently described and classified into 33 genera, twelve families and four orders (Redecker et al., 2013; Wijayawardene et al., 2018).

Molecular identification of AM fungi to genera and family is usually possible based on sequences of small subunit (SSU) or large subunit (LSU) regions of rDNA genes (Redecker et al., 2013; Öpik et al., 2014). However, species level inference based on rDNA genes is difficult due to high levels of intra species variation (Stockinger et al., 2009; House et al., 2016). While the rDNA operon is commonly found in a multi-copy tandem repeat organization across fungi (Lofgren et al., 2019), in AM fungi different rDNA variants can be scattered across the genome (VanKuren et al., 2013; Maeda et al., 2018) and lack the usual tandem organization (Maeda et al., 2018). The fact that rDNA genes are present as paralogs in AM fungal genomes likely explains the high levels of within strain and species diversity of rDNA sequences.

Limitations of single-locus phylogenetic inference and paralogous nature of rDNA genes in AM fungi calls for the need to generate extensive ortholog datasets from taxa representing different families, in order to accurately infer phylogenetic relationships among AM fungal lineages. One approach in this direction was achieved in a recent study using spore transcriptomic data for phylogenomic analysis of nine taxa from seven families (Beaudet et al., 2018). In this study, Glomerales was recovered as polyphyletic, in contrast to earlier rDNA phylogenies where Glomerales was found to be monophyletic (Krüger et al., 2012). Other phylogenomic studies have not addressed relations among families largely due to limited taxon sampling (Morin et al., 2019; Sun et al., 2019; Venice et al., 2020; Li et al., 2021). Due to difficulties in obtaining enough pure DNA for whole genome sequencing, available genomic data still represent only a fraction of the diversity of AM fungi.

Progress in AM fungal genomics has been limited by their biology. Arbuscular mycorrhizal fungi complete their life cycle

underground, as obligate symbionts of plant roots, and reproduce through multinuclear asexual spores (Bonfante and Genre, 2010). The spores are the largest isolable structure produced, but large-scale isolation of spores is needed in order to obtain enough DNA for whole genome sequencing. Such large-scale harvest of spores is only possible with AM fungi grown in axenic cultures where the fungus produces spores in a compartment separate from the transformed plant roots that it associates with (Tisserant et al., 2013), or for the rare taxa, such as *Diversispora epigaea* that forms fruitbodies above ground and from which large amounts of spores can be extracted (Sun et al., 2019). Axenic culturing methods are time-consuming and have only been successful for a handful of species (Kameoka et al., 2019). Due to the difficulty of producing clean cultures and isolate high quality DNA extracts for the majority of AM fungal species, it has been a slow path toward genomic studies of AM fungi (Tisserant et al., 2013; Lin et al., 2014; Beaudet et al., 2018; Kobayashi et al., 2018; Morin et al., 2019; Sun et al., 2019; Venice et al., 2020).

AM fungal hyphae and their asexual spores are coenocytic, and sequence analysis of individual nuclei have been used to analyze intra organismal polymorphism mainly by mapping reads of single nuclei to reference genomes of the model AM fungus *Rhizophagus irregularis* (Lin et al., 2014; Ropars et al., 2016; Chen et al., 2018). To circumvent the obstacle of pure culturing, we recently presented a workflow that takes advantage of automated nuclei sorting by extracting nuclei from ungerminated spores, directly extracted from soil (Montoliu-Nerin et al., 2020). Single nuclei of *Claroideoglomus claroideum* were sorted, followed by whole genome amplification (WGA) and sequencing. Finally, the data from several nuclei were combined to build a *de novo* genome assembly. With this novel workflow AM fungal genome assemblies can be obtained from as little as one single spore, independently of the species ability to grow in axenic cultures (Montoliu-Nerin et al., 2020). Similar approaches have been successfully applied in other organisms for which limited access to pure biological material suitable for extraction of high-quality DNA has prevented genome sequencing (Stepanauskas and Sieracki, 2007; Woyke et al., 2009; Heywood et al., 2011; Yoon et al., 2011; Walker et al., 2014; Wideman et al., 2019).

In this study, we sorted and sequenced nuclei from AM fungal spores representing species across Glomeromycota, aiming to obtain genomic information from two taxa for each genus. To evaluate the quality of assemblies generated by our workflow, we included *Rh. irregularis* DAOM197198, a strain for which a reference genome generated from an axenic culture is available (Chen et al., 2018), and compared this published assembly with our newly generated assembly. A final count of 21 strains, from 12 genera, across seven families were successfully sequenced, and *de novo* genome assemblies were constructed. Our dataset includes 15 species for which genome data was not previously available. This comprehensive taxon sampling allowed us to infer evolutionary relationships among AM fungi. Furthermore, the release of new whole genome assemblies provides a resource to the research community, for those interested in further exploring genetics and evolution of this important group of fungi.

MATERIALS AND METHODS

Fungal Strains

Taxa were initially selected to represent 15 genera across eight families in Glomeromycota (Schüßler and Walker, 2010; Redecker et al., 2013), aiming for two species per genus (**Supplementary Table 1**). The isolates were obtained as whole inoculum from the International culture collection of (vesicular) arbuscular mycorrhizal fungi (INVAM) at West Virginia University, Morgantown, WV, USA, or James D. Bever's lab, University of Kansas, USA, with the exception of *Rh. irregularis* DAOM197198 which was obtained as a tube of spores from Agriculture and Agri-food Canada, Government of Canada. In addition to the AM fungi sampled for this study, we included published annotated genome- and transcriptome assemblies of AM fungi (Glomeromycota; 13 taxa, **Supplementary Table 2**). Furthermore, we included all species of the closest sister lineages with available genome assemblies and annotations (December 2019) in the JGI (Joint Genome Institute) database, i.e. Mortierellomycota (2 taxa, Mondo et al., 2017a; Uehling et al., 2017) and Mucoromycota (12 taxa, Ma et al., 2009; Wang et al., 2013; Schwartze et al., 2014; Chibucos et al., 2016; Corrochano et al., 2016; Mondo et al., 2017a,b; Chang et al., 2019) (**Supplementary Table 2**). Finally, members of Dikarya were included as outgroup, with three representatives of Ascomycota, one taxon each from the subphyla Taphrinomycotina (Pomraning et al., 2018), Saccharomycotina (Wood et al., 2002), and Pezizomycotina (Martin et al., 2010); and three representatives of Basidiomycota, one taxon each from the subphyla Agaricomycotina (Martin et al., 2008), Ustilaginomycotina (Kämper et al., 2006), and Pucciniomycotina (Schwessinger et al., 2018) (**Supplementary Table 2**).

Nuclear Sorting and Whole Genome Amplification

Spores were extracted from whole inoculum cultures by sieving, followed by a sucrose gradient centrifugation as described in Montoliu-Nerin et al. (2020). A single spore or a pool of spores (**Supplementary Table 1**) were then rinsed and stored in 20 μ l ddH₂O in a 1.5 ml tube. After adding 180 μ l of 1× PBS spores were crushed with a sterile pestle and DNA was stained by adding 1 μ l of 200× SYBR Green I Nucleic Acid stain (InvitrogenTM, Thermo Fisher Scientific, MA, USA). Sorting of the nuclei proved to be more successful when the crushed spore solution was transferred to the small 0.5 ml tube for staining. This allowed the spore debris to settle while the nuclei remained in solution. The sample was left staining for 30–60 min, and lower sorting performance was observed when exceeding that time. The nuclear sorting was performed at the SciLifeLab Microbial Single Cell Genomics Facility with a MoFloTM Astrios EQ sorter (Beckman Coulter, USA), as in Montoliu-Nerin et al. (2020). Briefly, a 100 μ m nozzle was used and the sheath fluid, 0.1 μ m filtered 1× PBS, was run at 25 psi. Nuclei populations were identified via nuclei acid staining using the 488 nm laser and a 530/40 nm bandpass filter over forward or side scatter. Individual nuclei were deposited into 96- or 384-well plates using stringent single-cell sort settings (single

mode, drop envelope 1). These sort-settings abort target cells if another particle of any type is in the same or the neighboring drop, thereby increasing the number of abortions while ensuring that only one particle gets sorted per well. Each day of sorting, the sort precision was determined with beads sorted onto a slide and counted manually under the microscope. A low event rate was used to decrease the risk of sorting doublets, for most samples below 500 events per second with a drop generation of >40,000 per second corresponding to well-below 1% of nuclei in the samples. Most of the remaining particles were low in SYBR Green fluorescence. To each plate, 48 wells were used for sorting single nuclei or up to four pools of five nuclei, leaving the rest of the wells empty. Plates with sorted nuclei were stored at -80°C .

DNA from the nuclei samples was amplified with the enzyme Phi29 via multiple displacement amplification (MDA) under clean (i.e., amplicon and contaminant free) conditions using the RepliPhi kit (Epicenter) in a 15 μ l reaction volume in 96-well plates or with the Repli-g Single Cell kit (Qiagen) in a 10 μ l reaction volume in 384-well plates. The nucleic acid stain SYTO 13 was added to the reaction to follow the DNA amplification over time. Protocol including plate size and MDA kit was changed over time (**Supplementary Table 1**).

Sequencing of Amplified Nuclei Samples

We screened MDA nuclei samples by PCR amplification of rDNA markers using fungal and bacterial specific primers, following the protocol in Montoliu-Nerin et al. (2020). Multiple displacement amplification nuclei samples that scored positive for fungi and negative for bacteria were selected for sequencing. For samples with enough DNA the TruSeq PCRfree DNA library preparation kit (Illumina Inc.) was used. In total, 7–24 nuclei from each isolate were independently sequenced with Illumina HiSeq-X, at the SNPandSEQ Technology Platform in Uppsala at the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory, as in Montoliu-Nerin et al. (2020). Detailed information on sorting, MDA results, PCR screening, and selection of nuclei samples is available in **Supplementary Data File 1**.

Genome Assembly and Strain Verification

Whole genome assembly was performed according to assembly workflow 3 as described in Montoliu-Nerin et al. (2020), in which all sets of reads from individually sequenced nuclei samples from each strain were combined and normalized using bbnorm of BBMap v.38.08 (Bushnell, 2016), setting an average depth of 100X, and then assembled using SPAdes v.3.12.0 (Bankevich et al., 2012). We chose this workflow for the current study as it gives a good representation and accuracy of single copy genes, making it more suitable for downstream phylogenomic analyses than the other two workflows developed (Montoliu-Nerin et al., 2020). We used Quast v.4.5.4 (Gurevich et al., 2013) to quantitatively assess the assemblies (**Supplementary Table 3**) and ran BUSCO v.3.0.2b (Simão et al., 2015) to assess completeness of the genome, using fungi_odb9 as lineage setting, and rhizopus_oryzae

as species set (**Supplementary Tables 3, 4**). Raw reads and *de novo* genome assemblies are deposited in ENA in the project PRJEB45340.

To verify strain identity based on a reconstructed ribosomal gene phylogeny, we extracted the ribosomal gene operon from each newly assembled genome. For strains in the family Claroideoglomeraceae only one of its highly diverging rDNA sequences (VanKuren et al., 2013) was retrieved as a complete operon. In earlier work, when the genome of *C. claroideum* was assembled by combining single nuclei assemblies, we could identify both rDNA variants, but not when the genome assembly was generated from combined and normalized reads (Montoliu-Nerin et al., 2020). The SSU region was combined with the taxon rich SSU alignment from Krüger et al. (2012). The whole rDNA operons extracted from genome assemblies with verified identity, were aligned and a phylogeny was reconstructed with RAxML v.8.2.10 (Stamatakis, 2014), implementing the GTR model and with IQ-TREE v.1.6.5 (Nguyen et al., 2015), using ModelFinder (Kalyaanamoorthy et al., 2017) and searching for the best partitioning scheme. We ran both analyses with 1,000 bootstrap replicates. Extracted rDNA operons for all *de novo* genome assemblies available in the linked public OSF repository.

Genome Annotation

Each genome assembly was annotated using a snakemake workflow (Köster and Rahmann, 2012) v.2.0. The workflow is publicly available at https://bitbucket.org/scilifelabs/genemark_fungal_annotation/ (tag v.3.0, with minor updates providing the same functionality). Briefly, repeats and transposable elements were *de novo* predicted in each of the assemblies using RepeatModeler v.1.0.8 (Smit and Hubley, 2008) and the resulting repeat library was used to mask each genome assembly using RepeatMasker v.4.0.7 (Smit et al., 2015). UniProt/Swiss-Prot (Consortium, 2018) protein sequences (downloaded 8 May 2018) were aligned to each of the repeat-masked genome assemblies with MAKER v.3.01.1-beta (Cantarel et al., 2008). Protein coding genes were *de novo* predicted from each of the repeat-masked genome assemblies with GeneMark-ES v.4.33 (Ter-Hovhannisyan et al., 2008), providing the genomic locations of UniProt/Swiss-Prot proteins aligned to the genome assembly to guide the gene predictions. Minimum contig size to be included in self-training of the GeneMark gene prediction algorithm was calculated to include at least 10 Mb of training data, depending on the level of fragmentation of the assembly, and was set accordingly using the parameter “-min_contig” (Table of specific parameter used for each assembly is available in the linked public OSF). Protein and gene names were assigned to the gene predictions using a BLASTp v.2.7.1 (Camacho et al., 2009) search of predicted protein sequences against the UniProt/Swiss-Prot database with default e-value parameters (1×10^{-5}). InterProScan v.5.30-69.0 (Cock et al., 2013) was used to collect predictive information about the predicted proteins' functions.

Assessing Assembly Quality Using *Rhizophagus irregularis* DAOM197198

To confirm the accuracy of assemblies generated in our workflow we included the reference strain *Rh. irregularis* DAOM197198 (**Supplementary Table 1**) and compared our *de novo* genome assembly to a published high-quality genome assembly DAOM197198 v.2.0 (**Supplementary Table 2**) (Chen et al., 2018). Including this well-characterized strain allowed us to assess the performance of our assembly workflow. To assess efficiency and coverage of single nuclei MDA and sequencing, we mapped reads from individual nuclei against the published reference assembly and to our *de novo* assembly of *Rh. irregularis* DAOM197198, using BWA 0.7.15 (Li and Durbin, 2009), and measured both % of reads mapping and % of assembly covered with mapped reads using Qualimap 2.2.1 (Okonechnikov et al., 2016) and bamtools v.2.3.0 stats (Barnett et al., 2011). We also tested for polymorphism introduced during MDA by pairwise alignment of the 271 BUSCO genes retrieved from both assemblies using MAFFT v.7.407 (Katoh and Standley, 2013). Percentage similarity for the alignments was calculated with eslistat in HMMer v.3.2.1 (Hancock and Bishop, 2004). Finally, to take advantage of the ready-made comparative analysis of OrthoFinder v.2.4.0 (Emms and Kelly, 2018), we used this software (with standard settings) to identify orthogroups in the two genome assemblies.

Phylogenomic Analyses

Phylogenomic analyses were performed at different taxonomic scales, using six datasets with different taxon sampling (**Supplementary Table 5**). The first dataset was designed to explore the relationship of Glomeromycota with its sister phyla Mucoromycota and Mortierellomycota, and included genome data from Dikarya as outgroup (**Supplementary Tables 2, 3**). Two more datasets were designed to explore the relationships within Glomeromycota, including as many AM fungal taxa as possible, as well as members of its sister phyla. One of them included assembled transcriptomic data for Glomeromycota (Beaudet et al., 2018), while the other included only genomic data. For each of these two latter datasets, single copy orthologs (SCOs) were identified from the gene predictions using OrthoMCL v.2.0.9 (Li et al., 2003) with default parameters, requiring that SCOs were present in >50% of the taxa. Three additional datasets were designed to further explore conflicting topologies within Glomeromycota. Two of the three, were taxon-rich, including all species in Glomeromycota with available genome data (27 taxa), for one we retrieved SCOs present in >50% of the taxa, while for the other we retrieved SCOs present in all 27 taxa. The last of the three Glomeromycota datasets was designed to evaluate whether the inclusion of assemblies with low BUSCO completeness (<80%) impacts the phylogenetic reconstruction. Thus, this dataset included at least one species from each available genus with >90% estimated BUSCO completeness, except for *Ra. fulgida* and *Ac. colombiana*, for which BUSCO completeness was estimated to 82 and 87%, respectively (**Supplementary Tables 3, 4**). Single copy orthologs

that were present in all 15 taxa were included in this dataset (**Supplementary Table 5**).

Amino acid sequences were aligned using MAFFT v.7.407 (Katoh and Standley, 2013). Poorly aligned regions were removed using trimAl v.1.4.1 (Capella-Gutiérrez et al., 2009) with a gap threshold of 0.1 (0.2 in the dataset with only 15 taxa selected, **Supplementary Table 4**). Individual SCO alignments were removed if shorter than 100 amino acids. Single copy ortholog alignments were used either separately, to produce individual gene trees, or concatenated, to produce best maximum likelihood (ML) trees. Individual SCO alignments were concatenated into a supermatrix using the script geneStitcher.py (Schluter, 2016), which also produces a gene partition file. Lists of SCOs used for phylogenomic inferences from each dataset and their corresponding gene annotations are available in the linked public OSF repository.

Phylogenetic inferences based on the concatenated sets of SCOs were performed using two ML methods. First, ML phylogenies were inferred using RAXML v.8.2.10 (Stamatakis, 2014), with 100 bootstrap replicates, and with a partitioned model that treated each SCO as a separate partition and implementing the PROTOGAMMAWAG model for all partitions. Secondly, ModelFinder (Kalyaanamoorthy et al., 2017) was run for every partition, and a best ML tree was generated with 100 bootstrap replicates in IQ-TREE v.1.6.5 (Nguyen et al., 2015). Topologies and support values from both ML inference methods were highly comparable, therefore, we only present the RAXML topologies but adding support values from the IQ-TREE analysis. Phylogenetic inferences were also performed with ASTRAL-III v.5.7.3 (Zhang et al., 2018), a method consistent with a multi-species coalescent model. For this, we used individual gene trees inferred with IQ-TREE using the automated detection for the best-fitting model (-MFP) and 100 bootstrap replicates. The topological robustness was evaluated with a multi-locus bootstrapping (MLBS) and local posterior probabilities (LPP). For the dataset including Glomeromycota and its sister phyla, a Bayesian phylogeny was inferred using Phylobayes (Lartillot et al., 2009), under the site-heterogeneous CAT+GTR+G4 model on a total alignment of 144,177 amino acids. Two chains were run and convergence was evaluated using the commands tracecomp and bpcomp in Phylobayes, which was achieved after 120,000 and 200,000 generations.

We evaluated the phylogenetic placement of Glomeromycota in relation to Mucoromycota and Mortierellomycota, and the relationships within Glomeromycota, specifically Claroideoglomeraceae, Glomeraceae and Diversisporales. For this, we examined the support among individual gene trees for alternative branching orders. We performed a polytomy test in ASTRAL-III v.5.7.3 to identify evidence for hard polytomies. The test uses quartet gene tree frequencies to evaluate whether a polytomy could be rejected (Sayyari and Mirarab, 2018).

To further explore the relationships among lineages of AM fungi, splits networks were produced for two datasets (**Supplementary Table 5**) using IQ-TREE v.1.6.5 (Nguyen et al., 2015) with the command `iqtree -net`. Networks were visualized in SplitsTree5 (Huson and Bryant, 2006) with a maximum

dimension of 2. For the dataset with 15 selected AM fungal taxa, topologies branching over the tree landscape were also visualized and the consensus topologies were analyzed using DensiTree v.2.01 (Bouckaert and Heled, 2014) based on the previously inferred individual gene trees from IQ-TREE.

RESULTS

Presenting 21 *de novo* Genome Assemblies of AM Fungi

Using our novel workflow for *de novo* assembly of genomes by combining single nuclei sequence data (Montoliu-Nerin et al., 2020), we aimed to sequence 31 AM fungal isolates, with at least two taxa from each 15 genera across eight families (**Supplementary Table 1**). Spores from all 31 isolates were extracted for nuclei sorting and DNA amplification. For two of the taxa, *Archaeospora trappei* and *Entrophospora infrequens*, we failed to sort nuclei, and these were thus omitted from subsequent methods. After WGA with MDA on the sorted nuclei, samples from the remaining 29 isolates were screened by PCR amplification of the rDNA barcode region for presence of DNA of fungal and bacterial origin. Presence of fungal DNA was confirmed for 25 of the isolates, while samples from four isolates did not amplify the fungal rDNA barcode region and were thus excluded from sequencing (**Supplementary Table 1**). Genome assemblies of the 25 isolates ranged from 50 to 493 Mb in size, with numbers of gene predictions ranging from 11,400 to 46,500 and BUSCO completeness between 55 and 95% (**Supplementary Tables 3, 4**). Four of these assemblies were later removed due to misidentification of strains, see Isolate Identification in rDNA-based Phylogeny below, resulting in a final number of 21 genomes presented and used in the phylogenomic analyses.

Based on the comparison of *Rh. irregularis* DAOM197198 genome assemblies, we found that single nuclei MDA and sequencing were highly accurate and efficient in our workflow. On average, around 99% of the reads mapped to both our *de novo* genome assembly and the published reference genome v.2.0 of *Rh. irregularis* DAOM197198 (**Supplementary Table 6**). Reads from individual nuclei covered on average 50% of both assemblies and when combined the reads covered close to 95% of the reference genome v.2.0 (**Supplementary Table 6**). Together these results demonstrate that reads from single amplified and sequenced nuclei fully match the published reference and that the whole genome is represented among the reads. Pair-wise alignment of the 271 BUSCO genes retrieved in both assemblies of *Rh. irregularis* DAOM197198 demonstrate high consistency with an average similarity of 99.7% across nucleotide alignments. Of the 271 pairwise aligned BUSCO genes, a total of 260 were identical between the two assemblies, corresponding to 96% of the retrieved BUSCO genes. However, only 60% similarity was detected in one of the 271 pairwise alignments, and ten alignments ranged in similarity between 84 and 99% (**Supplementary Data File 2**). High similarity in pairwise alignments of BUSCO genes retrieved from the two assemblies demonstrates that random errors possibly introduced

during MDA are not retained to a large extent in genes in the assembled genome when reads from single nuclei are combined and normalized before assembling with SPAdes. In our assembly of *Rh. irregularis* DAOM197198, 23,258 genes were predicted (**Supplementary Table 3**) compared to 26,183 genes predicted in the published assembly of *Rh. irregularis* DAOM197198 v.2.0 (Chen et al., 2018). We demonstrate that our *de novo* genome assembly for *Rh. irregularis* DAOM197198 contained a largely overlapping set of genes in orthogroups present in the published *Rh. irregularis* DAOM197198 reference genome v.2.0. Across the two genome assemblies of the same strain, a total of 13,908 orthogroups were identified including 88% of all predicted genes across the two assemblies, of these, 94% were shared between the two genome assemblies (**Supplementary Table 7**). Interestingly, both genome assemblies contain orthogroups not recovered in the other, 403 unique to v.2.0 and 380 unique to our *de novo* assembly (**Supplementary Table 7**).

Isolate Identification in rDNA-based Phylogeny

The complete rDNA operon, including SSU, ITS1, 5.8s, ITS2, and LSU regions was extracted from the 25 newly generated genome assemblies. To confirm genus level identity of the 25 isolates for which we generated genome assemblies in this study, the SSU rDNA region was extracted and placed into the taxon-rich Glomeromycota phylogeny of Krüger et al. (2012) (**Supplementary Figure 1**). For five isolates, the species name did not correspond with the phylogenetic placement, revealing that these isolates were originally misidentified. Four of these were removed. First, the isolate *Rhizophagus intraradices* FL208A clustered within the genus *Funnelformis* (**Supplementary Figure 1**), more specifically, together with samples of *Funnelformis mosseae*. Morphological examination of this strain was consistent with its original identification as *Rh. intraradices*. We could not verify that spores with the correct morphology had been extracted for nuclei sorting. Therefore, this genome assembly was excluded from further analyses. The isolate *Funnelformis caledonius* UK204 also clustered with samples representing *F. mosseae* (**Supplementary Figure 1**), but since the genus placement was correct the strain was kept as *F. caledonius* for further analysis. The isolate *Di. epigaea* AZ150B was phylogenetically misplaced based on its rDNA SSU sequence (**Supplementary Figure 1**), and the assembly had the highest GC content among our assemblies (**Supplementary Table 3**), probably due to bacterial contamination. The MDA success for this strain was low and this was the only strain for which we included two nuclei samples, out of seven sequenced, where PCR had indicated the presence of bacterial DNA. We thus decided to discard this assembly since a genome of *Di. epigaea* is publicly available (Sun et al., 2019) and was included in the analyses. Finally, the isolates *Archaeospora scheckii* CL383 and *Septoglomus viscosum* MD215 were placed in the family Paraglomeraceae (**Supplementary Figure 1**), and subsequently eliminated from further analyses, as two *Paraglomus* isolates were already included. After this confirmation step, *de novo* genome assemblies representing 21 isolates were kept for the

phylogenomic analyses. A phylogenetic analysis of the entire extracted rDNA operon from the 21 genome assemblies, as well as that of *C. claroideum* previously generated in our group, showed that, in line with earlier phylogenetic results based on rDNA genes (Redecker et al., 2013), Glomerales is monophyletic, albeit with bootstrap support (BS) of just over 80% (**Supplementary Figure 2**).

Phylogenomic Analysis of Glomeromycota

To place Glomeromycota in relation to its sister phyla, phylogenetic trees were built from a dataset that included members of Glomeromycota, Mucoromycota and Mortierellomycota, and Dikarya as outgroup, with 178 SCOs that were shared among >50% of the taxa. The concatenated alignment had a length of 76,737 amino acids. In the RAXML phylogeny, Glomeromycota and Mortierellomycota form a monophyletic clade (80% BS), with Mucoromycota as their sister group (100% BS) (**Supplementary Figures 3, 4A**). However, in the IQ-TREE phylogeny, Mortierellomycota and Mucoromycota form a monophyletic clade (43% BS), sister to Glomeromycota (100% BS) (**Supplementary Figure 4B**). The ASTRAL analysis recovered the same topology as the RAXML analysis, but with low support (LPP = 0.53) (**Supplementary Figure 4C**). The quartet gene tree frequencies were very similar for the three alternative topologies ($q_1 = 0.37$, $q_2 = 0.29$, $q_3 = 0.34$), suggesting that a polytomy cannot be rejected (p -value = 0.71) (**Supplementary Figure 5**). The relationships among the three sister phyla remain unresolved in our analysis, likely because of the incomplete taxon sampling, in particular for Mortierellomycota that was only represented by two species.

Two more datasets were designed to explore the relationships within Glomeromycota. One of them included published transcriptomic data from nine genera of AM fungi (Beaudet et al., 2018). However, only 17 SCOs shared among >50% of the taxa were retrieved (**Supplementary Table 5**). This low number is likely resulting from the fact that the transcriptomic dataset is less complete, ranging from 22 to 87% BUSCO completeness with an average of 57% across nine species (Beaudet et al., 2018), compared to the genomic data generated in our study which ranged from 44 to 95%, averaging at 83% across 21 strains included in the phylogenomic analysis. The phylogenetic placement of the strains with transcriptomic data is consistent with the placement of our newly sequenced strains (**Supplementary Figure 6**), but due to their low BUSCO values and little overlap of SCOs, the transcriptomic data was excluded from further analysis without decreasing the taxonomic breadth, while allowing us to work with a more comprehensive set of SCOs. The other dataset included published genome data and 21 newly sequenced strains of AM fungi, as well as representatives from Mortierellomycota and Mucoromycota. A concatenated alignment of 371 SCOs shared among >50% of the taxa produced a total alignment of 144,177 amino acids. All represented Glomeromycota families form well-supported monophyletic lineages in both ML and ASTRAL analyses (**Figure 1**; **Supplementary Figure 7**). This supports available phylogenetic inferences based on a combination of morphology and rDNA data (Redecker et al., 2013). We found, however,

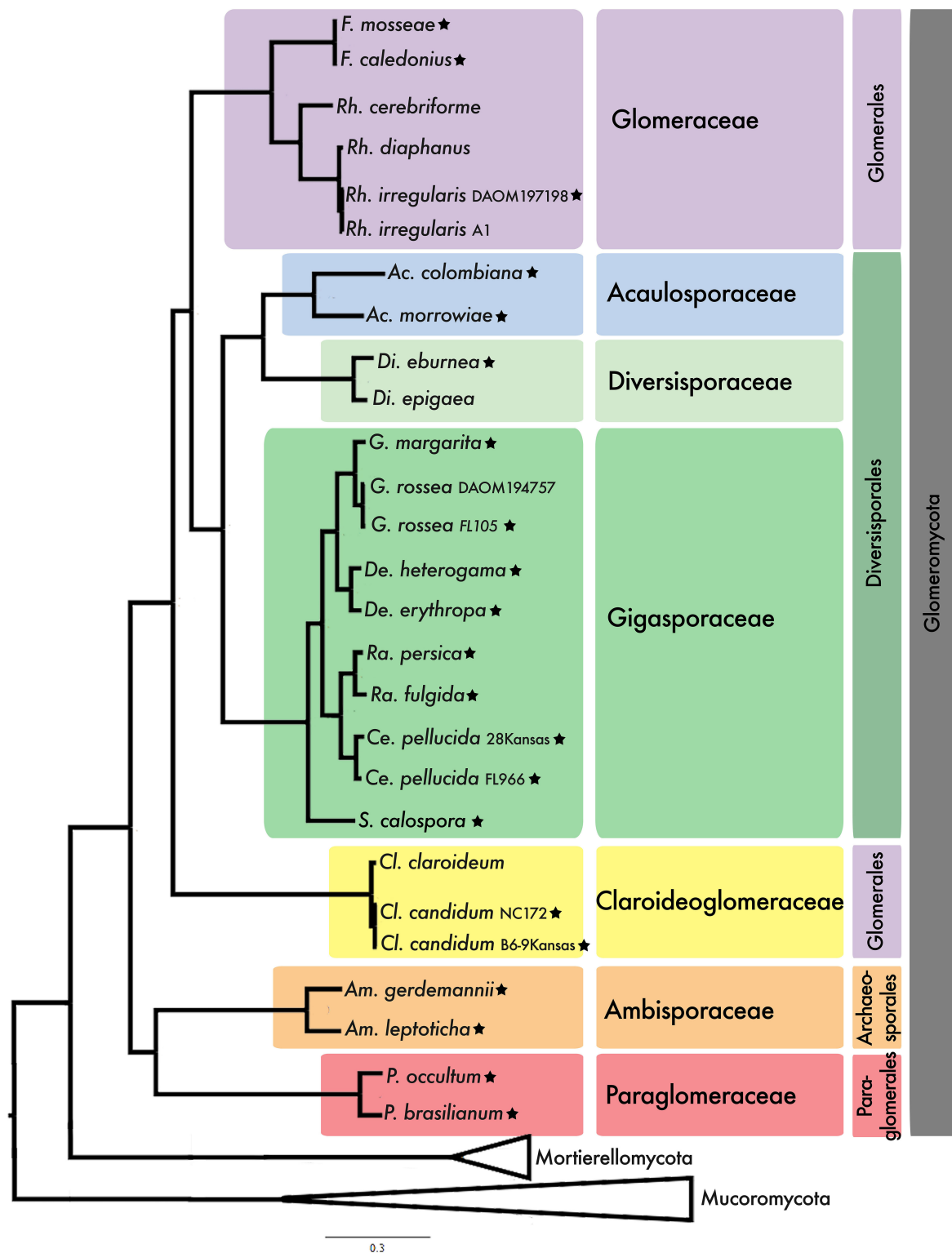


FIGURE 1 | Best maximum likelihood tree inferred with RAxML from a concatenated alignment of 371 single copy orthologs shared among >50% of the taxa. The same topology was recovered using IQ-TREE and Bayesian inference. All nodes have a bootstrap support value of 100 in both analyses, and posterior probabilities of 1. Mucoromycota was used as outgroup. Stars following the taxon name mark newly sequenced strains from this study. Current taxonomic assignment based on Redecker et al. (2013) is color coded, at the levels of family and order. Strain identifiers are included in the taxa label when more than one node has the same species name. See expanded tree in **Supplementary Figure 7**.

that the order Glomerales is polyphyletic, with Glomeraceae recovered as sister to the order Diversisporales (100% BS), while the family Claroideoglomeraceae forms a sister clade to the two (Figure 1; Supplementary Figure 7). The ASTRAL analysis recovered these same relationships but with low support (95% MLBS; 0.73 LPP) (Supplementary Figure 7). The ASTRAL analysis showed that the quartet gene tree frequencies of the three possible topologies were very similar ($q_1 = 0.37$, $q_2 = 0.33$, $q_3 = 0.30$), and a polytomy could not be rejected with this dataset (p -value = 0.31) (Supplementary Figure 8).

Exploring Conflicting Topologies

To further study contentious relationships within Glomeromycota (Claroideoglomeraceae, Glomeraceae, and Diversisporales), three datasets including only members of Glomeromycota, but with different taxon and/or gene sampling were produced (Supplementary Table 5). The first included 27 taxa with 1,737 SCOs present in >50% of the taxa (Supplementary Tables 2, 3). This dataset produced a concatenated alignment of 702,801 amino acids. The second included the same taxa, but only the 31 SCOs present in all taxa and produced a concatenated alignment of 15,443 amino acids. The third dataset included a selection of 15 *de novo* assembled Glomeromycota genomes with the highest quality (Supplementary Tables 3, 4) that represent all families with at least one species per genus. This last dataset was used to obtain a greater number of SCOs shared among all taxa. It included 799 SCOs, which resulted in a concatenated alignment of 476,329 amino acids.

Claroideoglomeraceae was well-supported as sister to Glomeraceae and Diversisporales in both the ML (100% BS) and ASTRAL (100% MLBS; 1.0 LPP) phylogenies, with the datasets that included 1,727 and 799 SCOs (Supplementary Figures 9, 10). The quartet-based analyses supported this branching ($q_1 = 0.4$; $q_2 = 0.33$; $q_3 = 0.27$), for both datasets (Figures 2A,C), and a polytomy was rejected (p -value = 0). The same topology was weakly recovered in the ASTRAL analysis of the dataset with 31 SCOs (36% MLBS; 0.54 LPP) (Supplementary Figure 11), but the ML analysis inferred Claroideoglomeraceae as sister to Diversisporales with weak support (59% BS) (Supplementary Figure 11A). The quartet gene tree frequencies favored Glomeraceae as sister to Diversisporales ($q_1 = 0.39$, $q_2 = 0.26$, $q_3 = 0.34$), but a polytomy could not be rejected with this dataset (p -value = 0.66) (Figure 2B). One of the alternative topologies that is recovered by rDNA genes (Supplementary Figures 1, 2), where Claroideoglomeraceae is sister to Glomeraceae in a monophyletic Glomerales, is frequently recovered but never statistically supported in any of the three datasets (Figure 2).

In addition to the three possible topologies discussed above, there is a multitude of rare topologies among all 15 taxa, across the 799 single gene trees for SCOs shared among all taxa. These topologies are visualized using DensiTree (Bouckaert and Heled, 2014), in which the single gene trees are stacked on top of each other (Supplementary Figure 12). DensiTree shows that most genes support the topology, in which we recovered Glomeraceae as a sister group of Diversisporales,

followed by the topology in which Glomerales is recovered as a monophyletic clade (Supplementary Figure 12), which is consistent with the quartet-based analysis (Figure 2). Across all analyses described above, we found consistent support for the polyphyly of Glomerales and a new hypothesis for the evolutionary relationships among families of Glomeromycota (Figures 1, 2). However, a phylogenomic network of the datasets with 1,737 and 799 SCOs revealed a clear reticulation at the base of the tree, indicating that the early evolutionary relationships cannot be resolved with the available data (Figure 3; Supplementary Figures 13, 14). Future studies may shed light on the processes behind these relationships.

DISCUSSION

Glomeromycota encompass all known AM fungi with their characteristic life cycle involving an obligate association with plants (Bonfante and Venice, 2020) as well as the exceptional fungal taxa *Geosiphon pyriformis* which forms a mutualistic symbiosis with the cyanobacteria *Nostoc punctiforme* (Malar et al., 2021). In the current study we present a four-fold increase in the number of AM fungal genomes available, which was achieved thanks to the development of a workflow for genome assembly from multiple individually amplified and sequenced nuclei (Montoliu-Nerin et al., 2020).

The current workflow for generating *de novo* reference genomes of AM fungi was developed by our team to circumvent the need for culturing AM fungi for genomic studies (Montoliu-Nerin et al., 2020). Read mapping of data from 24 individually amplified and sequenced *Rh. irregularis* DAOM197198 nuclei demonstrates near complete coverage of the published *Rh. irregularis* DAOM197198 v.2.0 reference genome (Supplementary Table 6), suggesting that separate amplification of multiple nuclei compensates for uneven amplification of individual nuclei. Consistent recovery of orthogroups in our *de novo* genome assembly of *Rh. irregularis* DAOM197198 (Supplementary Table 7) and evidence that mostly identical BUSCO genes are recovered from both assemblies provides further support that the presented workflow generates gene sequence data suitable for phylogenomic analysis. We anticipate that the release of these novel genome assemblies will become an important resource for the future study of AM fungi, supplementing already available AM fungal genomes (Tisserant et al., 2013; Lin et al., 2014; Chen et al., 2018; Kobayashi et al., 2018; Morin et al., 2019; Sun et al., 2019; Montoliu-Nerin et al., 2020).

Our phylogenomic analysis revealed a well-supported species tree for AM fungi. The relation of Glomeromycota to its two closest sister lineages, Mucoromycota and Mortierellomycota, had not yet been resolved with strong support, and based on previously available data the relation was best described as a polytomy (Li et al., 2021). Interestingly, with the addition of a considerable number of AM fungal genomes presented in this study we can still not reject a polytomy (Supplementary Figure 8). This further highlights the need for increased taxon sampling in the sister lineages of

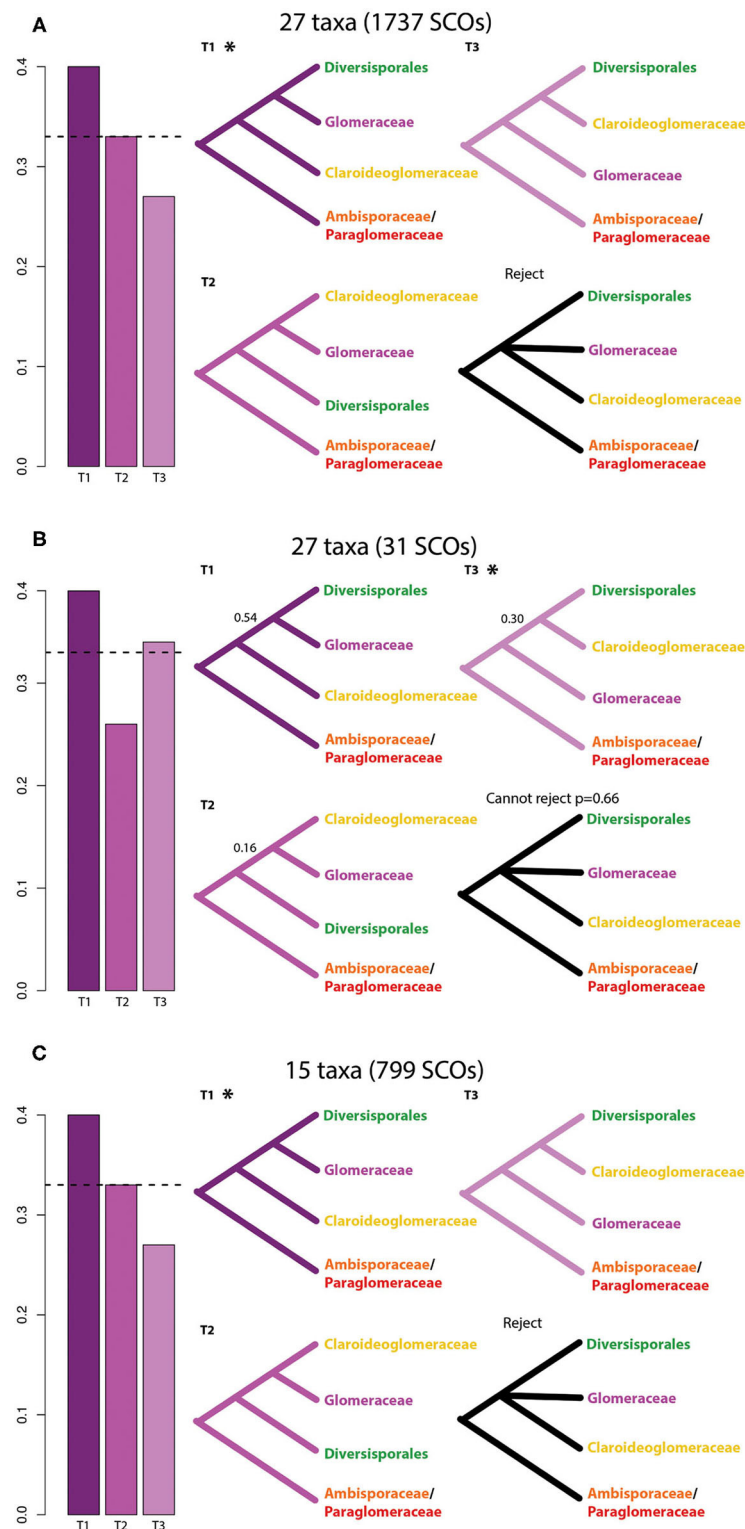
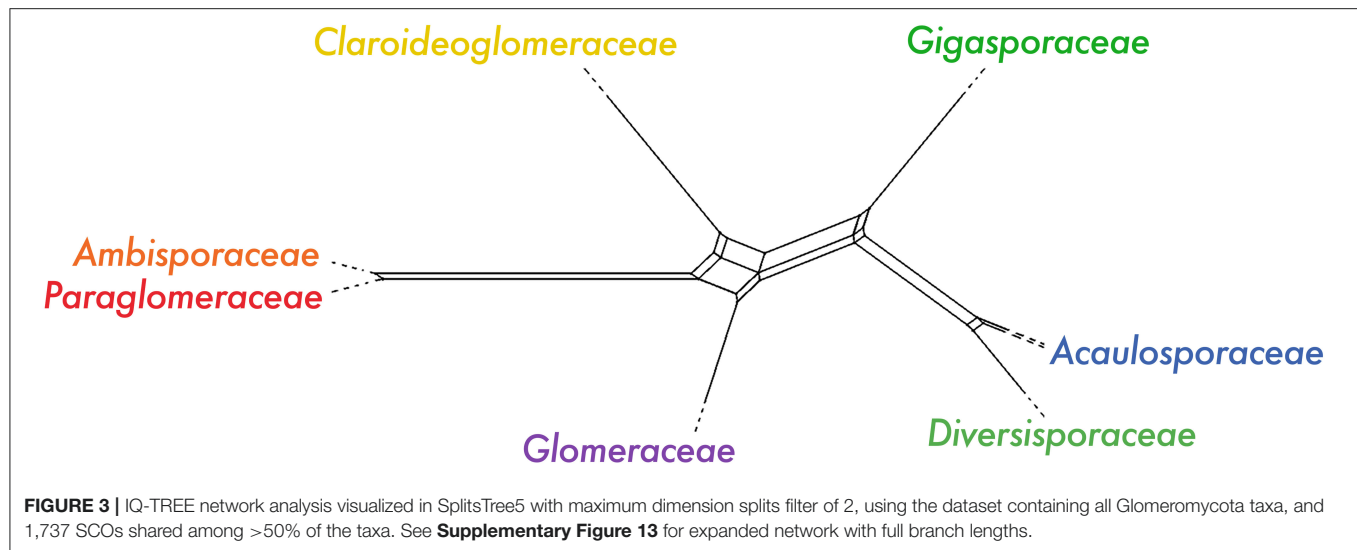


FIGURE 2 | Evaluation of support among individual gene trees for alternative hypotheses of the relationships within Glomeromycota based on three datasets. **(A)** Glomeromycota dataset with single copy orthologs (SCOs) that are present in >50% of the taxa (27 taxa/1737 SCOs). **(B)** Glomeromycota dataset including SCOs that are present in all the taxa (27 taxa/31 SCOs). **(C)** Glomeromycota dataset with a selection of 15 taxa (see methods and **Supplementary Table 4**) including SCOs that are present in all taxa. Bar graphs represent the gene tree quartet frequencies for three possible branching orders within Glomeromycota. T1 corresponds to the ASTRAL topology, T2 and T3 correspond to alternative topologies in ASTRAL. Dashed horizontal lines marked the expectation of a hard polytomy. The topologies inferred with the concatenation-based method (Maximum Likelihood) are marked with an asterisk (*). Local posterior probabilities are indicated only when below 1.0.



Glomeromycota, particularly in Mortierellomycota. Within Glomeromycota, we find that the seven family level lineages included in the analysis represent well-supported monophyletic lineages. Furthermore, while the order Diversisporales, including three families, was recovered as monophyletic we found that the order Glomerales with the two families Glomeraceae and Claroideoglomeraceae was not. Comprehensive phylogenetic studies with wide taxon sampling representing AM fungi have thus far mostly used rDNA sequences (Redecker et al., 2013) and recover Glomerales as monophyletic based on these markers. Similarly, our phylogenetic reconstructions using only the extracted rDNA operon from the *de novo* assembled genomes support Glomerales as monophyletic (**Supplementary Figure 2**). Glomerales was previously found to be polyphyletic in phylogenomic analyses using spore transcriptomic data from nine AM fungal species, where *Claroideoglomus* was recovered as sister to *Ambispora* and *Paraglomus* (Beaudet et al., 2018). In contrast to Beaudet et al. (2018), we recovered Claroideoglomeraceae as a sister to Glomeraceae and Diversisporales (**Figures 1, 2**). Previous phylogenomic studies using whole genomic data had not yet observed this topology due to limited taxon sampling (Morin et al., 2019).

The placement of Glomeraceae as a sister group of Diversisporales is well-supported in the phylogenies inferred using a concatenated dataset (**Figure 1**), as well as using a coalescence-based method (**Supplementary Figure 7**). Our analysis based on 27 Glomeromycota taxa and 1,737 SCOs that are present in >50% of the taxa strongly supported Claroideoglomeraceae as sister to Glomeraceae and Diversisporales. The dataset that included only 31 SCOs that are present in all taxa, showed low support for this relationship and a polytomy could not be rejected. These inconclusive results are most likely due to the small number of genes included in this dataset. It has been shown that phylogenetic inference can be robust to missing data (Wiens, 2003; Wiens and Morrill, 2011), therefore we expect that a more comprehensive set of

SCOs, even when not present in all taxa, will provide a more accurate phylogenetic reconstruction than a complete dataset representing few genes. However, by analyzing a dataset with the 15 best assemblies, including a representative of each genus, we demonstrated that the use of assemblies with low completeness (based on BUSCO values) does not impact the phylogenetic inference.

It is possible that the topological discordances are due to incomplete lineage sorting (Maddison and Knowles, 2006), caused by long coalescence times which complicates the assessment of an accurate evolutionary history. Different topologies could also result from gene flow among AM fungal lineages. Documented gene family expansions correlated with genome size in AM fungi (Tang et al., 2016), could distort phylogenetic histories since gene expansions and contractions can cause misidentification of SCOs, resulting in alignments between paralogs present as single copy with different evolutionary origins and histories. A better understanding on how variation in gene content and copy number variation influenced the different topologies could be achieved with a deeper phylogenetic study into the whole repertoire of paralogs, moving one step further from SCOs, which would also allow us to look more closely into the possible correlation between gene function and different evolutionary histories.

CONCLUSIONS

In the current study we present a considerable increase in the number of AM fungal genome assemblies available, thanks to the development of single nuclei sequencing and *de novo* assembling in AM fungi that we recently developed. As demonstrated for *Rh. irregularis*, variation in sequencing depth and coverage of single nuclei due to MDA, was accounted for in our *de novo* genome assemblies that provide a satisfactory representation of the genome content even when the assemblies generated are fragmented. Not all targeted species could be sorted, amplify or assemble equally well and species-specific

method development may be required for a more complete dataset. Nevertheless, we present a phylogenomic analysis of AM fungi based on the most comprehensive taxon sampling across Glomeromycota to date. Our results support current family-level classification and concur in one strongly supported topology. In this topology, the order Glomerales is polyphyletic, with the family Glomeraceae being recovered as a sister group to the order Diversisporales, with Claroideoglomeraceae as their sister group. The new genome data presented cover seven families of the phylum Glomeromycota and are expected to be a valuable contribution to the AM fungal research community.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ebi.ac.uk/ena>, PRJEB45340 and linked public OSF repository doi: 10.17605/OSF.IO/2ZYP4.

AUTHOR CONTRIBUTIONS

MM-N initiated the project together with AR and JB and developed the analysis together with MS-G and HJ. CB did the nuclei sorting and whole genome amplification. MM-N and MS-G performed the bioinformatic analyses. VK developed the annotation workflow which was ran by MM-N and MS-G. MM-N and MS-G wrote the manuscript with AR and HJ with input from all the authors. All authors contributed to the article and approved the submitted version.

REFERENCES

- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P., and Marth, G. T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* 27, 1691–1692. doi: 10.1093/bioinformatics/btr174
- Beaudet, D., Chen, E. C., Mathieu, S., Yildirim, G., Ndikumana, S., Dalpé, Y., et al. (2018). Ultra-low input transcriptomics reveal the spore functional content and phylogenetic affiliations of poorly studied arbuscular mycorrhizal fungi. *DNA Res.* 25, 217–227. doi: 10.1093/dnares/dsx051
- Bonfante, P., and Genre, A. (2010). Mechanisms underlying beneficial plant-fungus interactions in mycorrhizal symbiosis. *Nat. Commun.* 1, 1–11. doi: 10.1038/ncomms1046
- Bonfante, P., and Venice, F. (2020). Mucoromycota: going to the roots of plant-interacting fungi. *Fungal Biol. Rev.* 34, 100–113. doi: 10.1016/j.fbr.2019.12.003
- Bouckaert, R. R., and Heled, J. (2014). DensiTree 2: seeing trees through the forest. *bioRxiv*: 012401. doi: 10.1101/012401
- Bushnell, B. (2016). *BBMap Short Read Aligner*. Berkeley, CA: University of California.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*. doi: 10.1186/1471-2105-10-421

FUNDING

Funding for this project was provided by ERC (678792). VK was financially supported by the Knut and Alice Wallenberg Foundation as part of the National Bioinformatics Infrastructure Sweden at SciLifeLab. Open access funding provided by Uppsala University and funding for maintenance of AMF cultures by NSF 2027458.

ACKNOWLEDGMENTS

We thank Y. Strid and B. Ellis for assistance in the lab and J. Morton and W. Wheeler at INVAM culture collection. Nuclei sorting and whole genome amplification was done at the SciLifeLab Microbial Single Cell Genomics Facility at Uppsala University. Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. Computational analyses were performed on resources provided by SNIC through UPPMAX. The paper was greatly improved thanks to constructive comments from reviewers in Frontiers and in an earlier submission.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/ffunb.2021.716385/full#supplementary-material>

- Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., et al. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* 18, 188–196. doi: 10.1101/gr.6743907
- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Chang, Y., Desirò, A., Na, H., Sandor, L., Lipzen, A., Clum, A., et al. (2019). Phylogenomics of endogonaceae and evolution of mycorrhizas within mucoromycota. *New Phytol.* 222, 511–525. doi: 10.1111/nph.15613
- Chen, E. C., Morin, E., Beaudet, D., Noel, J., Yildirim, G., Ndikumana, S., et al. (2018). High intraspecific genome diversity in the model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*. *New Phytol.* 220, 1161–1171. doi: 10.1111/nph.14989
- Chibucos, M. C., Soliman, S., Gebremariam, T., Lee, H., Daugherty, S., Orvis, J., et al. (2016). An integrated genomic and transcriptomic survey of mucormycosis-causing fungi. *Nat. Commun.* 7, 1–11. doi: 10.1038/ncomms12218
- Cock, P. J., Grünig, B. A., Paszkiewicz, K., and Pritchard, L. (2013). Galaxy tools and workflows for sequence analysis with applications in molecular plant pathology. *PeerJ* 1: e167. doi: 10.7717/peerj.167
- Consortium, U. (2018). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 46, 2699. doi: 10.1093/nar/gky092
- Corrochano, L. M., Kuo, A., Marcet-Houben, M., Polaino, S., Salamov, A., Villalobos-Escobedo, J. M., et al. (2016). Expansion of signal transduction

- pathways in fungi by extensive genome duplication. *Curr. Biol.* 26, 1577–1584. doi: 10.1016/j.cub.2016.04.038
- Emms, D., and Kelly, S. (2018). OrthoFinder2: fast and accurate phylogenomic orthology analysis from gene sequences. *bioRxiv*: 466201.
- Gurevich, A., Savelyev, V., Vyahhi, N., and Tesler, G. (2013). QAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086
- Hancock, J. M., and Bishop, M. J. (2004). “HMMer,” in *Dictionary of Bioinformatics and Computational Biology*, eds J. M. Hancock and Marketa J. Zvelebil (John Wiley & Sons, Inc.). doi: 10.1002/9780471650126.dob0323.pub2
- Heywood, J. L., Sieracki, M. E., Bellows, W., Poulton, N. J., and Stepanauskas, R. (2011). Capturing diversity of marine heterotrophic protists: one cell at a time. *ISME J.* 5, 674–684. doi: 10.1038/ismej.2010.155
- Hibbett, D. S., Binder, M., Bischoff, J. F., Blackwell, M., Cannon, P. F., Eriksson, O. E., et al. (2007). A higher-level phylogenetic classification of the Fungi. *Mycol. Res.* 111, 509–547. doi: 10.1016/j.mycres.2007.03.004
- Hoeksema, J. D., Bever, J. D., Chakraborty, S., Chaudhary, V. B., Gardes, M., Gehring, C. A., et al. (2018). Evolutionary history of plant hosts and fungal symbionts predicts the strength of mycorrhizal mutualism. *Commun. Biol.* 1, 1–10. doi: 10.1038/s42003-018-0120-9
- House, G. L., Ekanayake, S., Ruan, Y., Schutte, U. M. E., Kaonongbua, W., Fox, G., et al. (2016). Phylogenetically structured differences in rRNA gene sequence variation among species of arbuscular mycorrhizal fungi and their implications for sequence clustering. *Appl. Environ. Microbiol.* 82, 4921–4930. doi: 10.1128/AEM.00816-16
- Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267. doi: 10.1093/molbev/msj030
- James, T. Y., Stajich, J. E., Hittinger, C. T., and Rokas, A. (2020). Toward a fully resolved fungal tree of life. *Annu. Rev. Microbiol.* 74. doi: 10.1146/annurev-micro-022020-051835
- Kalyanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285
- Kameoka, H., Tsutsui, I., Saito, K., Kikuchi, Y., Handa, Y., Ezawa, T., et al. (2019). Stimulation of symbiotic sporulation in arbuscular mycorrhizal fungi by fatty acids. *Nat. Microbiol.* 4, 1654–1660. doi: 10.1038/s41564-019-0485-7
- Kämper, J., Kahmann, R., Bölker, M., Ma, L.-J., Brefort, T., Saville, B. J., et al. (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* 444, 97–101. doi: 10.1038/nature05248
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kobayashi, Y., Maeda, T., Yamaguchi, K., Kameoka, H., Tanaka, S., Ezawa, T., et al. (2018). The genome of *Rhizophagus clarus* HR1 reveals a common genetic basis for auxotrophy among arbuscular mycorrhizal fungi. *BMC Genomics* 19, 1–11. doi: 10.1186/s12864-018-4853-0
- Köster, J., and Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* 28, 2520–2522. doi: 10.1093/bioinformatics/bts480
- Kozioł, L., Schultz, P. A., House, G. L., Bauer, J. T., Middleton, E. L., and Bever, J. D. (2018). The plant microbiome and native plant restoration: the example of native mycorrhizal fungi. *Bioscience* 68, 996–1006. doi: 10.1093/biosci/biy125
- Krüger, M., Krüger, C., Walker, C., Stockinger, H., and Schüßler, A. (2012). Phylogenetic reference data for systematics and phylotaxonomy of arbuscular mycorrhizal fungi from phylum to species level. *New Phytol.* 193, 970–984. doi: 10.1111/j.1469-8137.2011.03962.x
- Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25, 2286–2288. doi: 10.1093/bioinformatics/btp368
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503
- Li, Y., Steenwyk, J. L., Chang, Y., Wang, Y., James, T. Y., Stajich, J. E., et al. (2021). A genome-scale phylogeny of the kingdom Fungi. *Curr. Biol.* 31, 1653.e5–1665.e5. doi: 10.1016/j.cub.2021.01.074
- Lin, K., Limpens, E., Zhang, Z. H., Ivanov, S., Saunders, D. G. O., Mu, D. S., et al. (2014). Single nucleus genome sequencing reveals high similarity among nuclei of an endomycorrhizal fungus. *PLoS Genet.* 10, e1004078. doi: 10.1371/journal.pgen.1004078
- Lofgren, L. A., Uehling, J. K., Branco, S., Bruns, T. D., Martin, F., and Kennedy, P. G. (2019). Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles. *Mol. Ecol.* 28, 721–730. doi: 10.1111/mec.14995
- Ma, L.-J., Ibrahim, A. S., Skory, C., Grabherr, M. G., Burger, G., Butler, M., et al. (2009). Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. *PLoS Genet.* 5, e1000549. doi: 10.1371/journal.pgen.1000549
- Maddison, W. P., and Knowles, L. L. (2006). Inferring phylogeny despite incomplete lineage sorting. *Syst. Biol.* 55, 21–30. doi: 10.1080/10635150500354928
- Maeda, T., Kobayashi, Y., Kameoka, H., Okuma, N., Takeda, N., Yamaguchi, K., et al. (2018). Evidence of non-tandemly repeated rDNAs and their intragenomic heterogeneity in *Rhizophagus irregularis*. *Commun. Biol.* 1, 1–13. doi: 10.1038/s42003-018-0094-7
- Malar, M., Krüger, M., Krüger, C., Wang, Y., Stajich, J. E., Keller, J., et al. (2021). The genome of *Geosiphon pyriformis* reveals ancestral traits linked to the emergence of the arbuscular mycorrhizal symbiosis. *Curr. Biol.* 31, 1570.e4–1577.e4. doi: 10.1016/j.cub.2021.03.032
- Martin, F., Aerts, A., Ahrén, D., Brun, A., Danchin, E., Duchaussoy, F., et al. (2008). The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* 452, 88–92. doi: 10.1038/nature06556
- Martin, F., Kohler, A., Murat, C., Balestrini, R., Coutinho, P. M., Jaillon, O., et al. (2010). Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* 464, 1033–1038. doi: 10.1038/nature08867
- Mondo, S. J., Dannebaum, R. O., Kuo, R. C., Louie, K. B., Bewick, A. J., LaButti, K., et al. (2017a). Widespread adenine N6-methylation of active genes in fungi. *Nat. Genet.* 49, 964–968. doi: 10.1038/ng.3859
- Mondo, S. J., Lastovetsky, O. A., Gaspar, M. L., Schwardt, N. H., Barber, C. C., Riley, R., et al. (2017b). Bacterial endosymbionts influence host sexuality and reveal reproductive genes of early divergent fungi. *Nat. Commun.* 8, 1–9. doi: 10.1038/s41467-017-02052-8
- Montoliu-Nerin, M., Sánchez-García, M., Bergin, C., Grabherr, M., Ellis, B., Kutschera, V. E., et al. (2020). Building *de novo* reference genome assemblies of complex eukaryotic microorganisms from single nuclei. *Sci. Rep.* 10, 1–10. doi: 10.1038/s41598-020-58025-3
- Morin, E., Miyauchi, S., San Clemente, H., Chen, E. C., Pelin, A., de la Providencia, I., et al. (2019). Comparative genomics of *Rhizophagus irregularis*, *R. cerebriforme*, *R. diaphanus* and *Gigaspora rosea* highlights specific genetic features in Glomeromycota. *New Phytol.* 222, 1584–1598. doi: 10.1111/nph.15687
- Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Okonechnikov, K., Conesa, A., and García-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32, 292–294. doi: 10.1093/bioinformatics/btv566
- Öpik, M., Davison, J., and Moora, M., Zobel, M. (2014). DNA-based detection and identification of Glomeromycota: the virtual taxonomy of environmental sequences. *Botany* 92, 135–147. doi: 10.1139/cjb-2013-0110
- Parniske, M. (2008). Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nat. Rev. Microbiol.* 6, 763–775. doi: 10.1038/nrmicro1987
- Pomraning, K. R., Bredeweg, E. L., Kerkhoven, E. J., Barry, K., Haridas, S., Hundley, H., et al. (2018). Regulation of yeast-to-hyphae transition in *Yarrowia lipolytica*. *MSphere* 3, e00541-18. doi: 10.1128/mSphere.00541-18
- Redecker, D., Schüßler, A., Stockinger, H., Stürmer, S. L., Morton, J. B., and Walker, C. (2013). An evidence-based consensus for the classification of arbuscular mycorrhizal fungi (Glomeromycota). *Mycorrhiza* 23, 515–531. doi: 10.1007/s00572-013-0486-y

- Ropars, J., Toro, K. S., Noel, J., Pelin, A., Charron, P., Farinelli, L., et al. (2016). Evidence for the sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. *Nat. Microbiol.* 1, 16033. doi: 10.1038/nmicrobiol.2016.33
- Sayyari, E., and Mirarab, S. (2018). Testing for polytomies in phylogenetic species trees using quartet frequencies. *Genes* 9, 132. doi: 10.3390/genes9030132
- Schluter, D. (2016). *Phylogenetic Tools Phylogenetic Trees*. Available online at: <https://github.com/ballesterus/Utensils> (accessed June 01, 2019)
- Schüßler, A., Schwarzott, D., and Walker, C. (2001). A new fungal phylum, the Glomeromycota: phylogeny and evolution. *Mycol. Res.* 105, 1413–1421. doi: 10.1017/S0953756201005196
- Schüßler, A., and Walker, C. (2010). *The Glomeromycota: A Species List With New Families and New Genera*. The Royal Botanic Garden Kew, Botanische Staatssammlung Munich, and Oregon State University 19.
- Schwartz, V. U., Winter, S., Shelest, E., Marcet-Houben, M., Horn, F., Wehner, S., et al. (2014). Gene expansion shapes genome architecture in the human pathogen *Lichtheimia corymbifera*: an evolutionary genomics analysis in the ancient terrestrial mucorales (Mucoromycotina). *PLoS Genet.* 10, e1004496. doi: 10.1371/journal.pgen.1004496
- Schwessinger, B., Sperschneider, J., Cuddy, W. S., Garnica, D. P., Miller, M. E., Taylor, J. M., et al. (2018). A near-complete haplotype-phased genome of the dikaryotic wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici* reveals high interhaplotype diversity. *MBio* 9, e02275–17. doi: 10.1128/mbio.02275–17
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Smit, A., and Hubley, R. (2008). *2015 RepeatModeler Open-1.0*. Available online at: <http://www.repeatmasker.org> (accessed January 11, 2019).
- Smit, A., Hubley, R., and Green, P. (2015). *RepeatMasker Open-4.0. 2013–2015*. Available online at: <http://www.repeatmasker.org> (accessed January 11, 2019).
- Smith, S. E., and Read, D. J. (2010). *Mycorrhizal Symbiosis*. New York, NY: Academic Press.
- Spatafora, J. W., Chang, Y., Benny, G. L., Lazarus, K., Smith, M. E., Berbee, M. L., et al. (2016). A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* 108, 1028–1046. doi: 10.3852/16-042
- Stamatakis, A. (2014). RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Stepanaukas, R., and Sieracki, M. E. (2007). Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc. Nat. Acad. Sci. U.S.A.* 104, 9052–9057. doi: 10.1073/pnas.0700496104
- Stockinger, H., Walker, C., and Schüßler, A. (2009). 'Glomus intraradices' DAOM197198, a model fungus in arbuscular mycorrhiza research, is not *Glomus intraradices*. *New Phytol.* 183, 1176–1187. doi: 10.1111/j.1469-8137.2009.02874.x
- Sun, X., Chen, W., Ivanov, S., MacLean, A. M., Wight, H., Ramaraj, T., et al. (2019). Genome and evolution of the arbuscular mycorrhizal fungus *Diversispora epigaea* (formerly *Glomus versiforme*) and its bacterial endosymbionts. *New Phytol.* 221, 1556–1573. doi: 10.1111/nph.15472
- Tang, N., San Clemente, H., Roy, S., Bécard, G., Zhao, B., and Roux, C. (2016). A survey of the gene repertoire of *Gigaspora rosea* unravels conserved features among Glomeromycota for obligate biotrophy. *Front. Microbiol.* 7, 233. doi: 10.3389/fmicb.2016.00233
- Tedersoo, L., Sánchez-Ramírez, S., Koljalg, U., Bahram, M., Döring, M., Schigel, D., et al. (2018). High-level classification of the Fungi and a tool for evolutionary ecological analyses. *Fungal Divers.* 90, 135–159. doi: 10.1007/s13225-018-0401-0
- Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O., and Borodovsky, M. (2008). Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.* 18, 1979–1990. doi: 10.1101/gr.081612.108
- Tisserant, E., Malbreil, M., Kuo, A., Kohler, A., Symeonidi, A., Balestrini, R., et al. (2013). Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20117–20122. doi: 10.1073/pnas.1313452110
- Uehling, J., Gryganskiy, A., Hameed, K., Tschaplinski, T., Misztal, P., Wu, S., et al. (2017). Comparative genomics of *Mortierella elongata* and its bacterial endosymbiont *Mycoavidus cysteinexigens*. *Environ. Microbiol.* 19, 2964–2983. doi: 10.1111/1462-2920.13669
- VanKuren, N. W., den Bakker, H. C., Morton, J. B., and Pawlowska, T. E. (2013). Ribosomal RNA gene diversity, effective population size, and evolutionary longevity in asexual Glomeromycota. *Evol. Int. J. Organ. Evol.* 67, 207–224. doi: 10.1111/j.1558-5646.2012.01747.x
- Venice, F., Ghignone, S., Salvioli di Fossalunga, A., Amselem, J., Novero, M., Xianan, X., et al. (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, 122–141. doi: 10.1111/1462-2920.14827
- Walker, A. W., Duncan, S. H., Louis, P., and Flint, H. J. (2014). Phylogeny, culturing, and metagenomics of the human gut microbiota. *Trends Microbiol.* 22, 267–274. doi: 10.1016/j.tim.2014.03.001
- Wang, D., Wu, R., Xu, Y., and Li, M. (2013). Draft genome sequence of *Rhizopus chinensis* CCTCCM201021, used for brewing traditional Chinese alcoholic beverages. *Genome Announc.* 1, e0019512. doi: 10.1128/genomeA.00195-12
- Wideman, J. G., Lax, G., Leonard, G., Milner, D. S., Rodríguez-Martínez, R., Simpson, A. G., et al. (2019). A single-cell genome reveals diplonemid-like ancestry of kinetoplastid mitochondrial gene structure. *Philos. Trans. R. Soc. B* 374, 20190100. doi: 10.1098/rstb.2019.0100
- Wiens, J. J. (2003). Missing data, incomplete taxa, and phylogenetic accuracy. *Syst. Biol.* 52, 528–538. doi: 10.1080/10635150390218330
- Wiens, J. J., and Morrill, M. C. (2011). Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. *Syst. Biol.* 60, 719–731. doi: 10.1093/sysbio/syr025
- Wijayawardene, N. N., Pawlowska, J., Letcher, P. M., Kirk, P. M., Humber, R. A., Schüßler, A., et al. (2018). Notes for genera: basal clades of Fungi (including Aphelidiomycota, Basidiobolomycota, Blastocladiomycota, Calcarisporiellomycota, Caulochytriomycota, Chytridiomycota, Entomophthoromycota, Glomeromycota, Kickxellomycota, Monoblepharomycota, Mortierellomycota, Mucoromycota, Neocallimastigomycota, Olpidiomycota, Rozellomycota and Zoopagomycota). *Fungal Divers.* 92, 43–129. doi: 10.1007/s13225-018-0409-5
- Wood, V., Gwilliam, R., Rajandream, M.-A., Lyne, M., Lyne, R., Stewart, A., et al. (2002). The genome sequence of *Schizosaccharomyces pombe*. *Nature* 415, 871–880. doi: 10.1038/nature724
- Woyke, T., Xie, G., Copeland, A., Gonzalez, J. M., Han, C., Kiss, H., et al. (2009). Assembling the marine metagenome, one cell at a time. *PLoS ONE* 4, e5299. doi: 10.1371/journal.pone.0005299
- Yoon, H. S., Price, D. C., Stepanaukas, R., Rajah, V. D., Sieracki, M. E., Wilson, W. H., et al. (2011). Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* 332, 714–717. doi: 10.1126/science.1203163
- Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19, 153. doi: 10.1186/s12859-018-2129-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Montoliu-Nerin, Sánchez-García, Bergin, Kutschera, Johannesson, Bever and Rosling. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.