New Phytologist

Research

Phylogenomics of Endogonaceae and evolution of mycorrhizas within Mucoromycota

Ying Chang[1], Alessandro Desirò[2], Hyunsoo Na[3], Laura Sandor[3], Anna Lipzen[3], Alicia Clum[3], Kerrie Barry[3], Igor V. Grigoriev[3,4], Francis M. Martin[5], Jason E. Stajich[6], Matthew E. Smith[7], Gregory Bonito[2] and Joseph W. Spatafora[1]

[1]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA; [2]Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI 48824, USA; [3]US Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA; [4]Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA 94720, USA; [5]Institut national de la recherche agronomique, Laboratoire d'excellence ARBRE, Centre INRA-Grand Est, Unité mixte de recherche Inra-Université de Lorraine "Interactions Arbres/Microorganismes", 54280 Champenoux, France; [6]Department of Microbiology and Plant Pathology, University of California, Riverside, CA 92521, USA; [7]Department of Plant Pathology, University of Florida, Gainesville, FL 32611, USA

Author for correspondence:
Ying Chang
Tel: +1 541 908 7198
Email: changyi@oregonstate.edu

Received: *9 June 2018*
Accepted: *29 October 2018*

New Phytologist (2019) **222:** 511–525
doi: 10.1111/nph.15613

Key words: ectomycorrhiza, Endogonaceae, Endogonales, Mucoromycota, *Mycoplasma*-related endobacteria (MRE), mycorrhizal, plant-cell-wall-degrading enzymes.

Summary

- Endogonales (Mucoromycotina), composed of Endogonaceae and Densosporaceae, is the only known non-Dikarya order with ectomycorrhizal members. They also form mycorrhizal-like association with some nonspermatophyte plants. It has been recently proposed that Endogonales were among the earliest mycorrhizal partners with land plants. It remains unknown whether Endogonales possess genomes with mycorrhizal-lifestyle signatures and whether Endogonales originated around the same time as land plants did.
- We sampled sporocarp tissue from four Endogonaceae collections and performed shotgun genome sequencing. After binning the metagenome data, we assembled and annotated the Endogonaceae genomes. We performed comparative analysis on plant-cell-wall-degrading enzymes (PCWDEs) and small secreted proteins (SSPs). We inferred phylogenetic placement of Endogonaceae and estimated the ages of Endogonaceae and Endogonales with expanded taxon sampling.
- Endogonaceae have large genomes with high repeat content, low diversity of PCWDEs, but without elevated SSP/secretome ratios. Dating analysis estimated that Endogonaceae originated in the Permian–Triassic boundary and Endogonales originated in the mid–late Silurian. *Mycoplasma*-related endobacterium sequences were identified in three Endogonaceae genomes.
- Endogonaceae genomes possess typical signatures of mycorrhizal lifestyle. The early origin of Endogonales suggests that the mycorrhizal association between Endogonales and plants might have played an important role during the colonization of land by plants.

Introduction

Mycorrhizas are one of the most common symbioses on Earth. It is estimated that *c.* 90% of terrestrial plant species form mycorrhizal symbioses with fungi (Bonfante & Genre, 2010; van der Heijden *et al.*, 2015; Brundrett & Tedersoo, 2018). Mycorrhizal fungi provide the host plant with numerous elemental nutrients (e.g. phosphorus, nitrogen (N), manganese, calcium) and enhanced water uptake. In return they receive carbon (C) from the host plant in the form of glucose, and thus are central players in nutrient cycling in terrestrial ecosystems (Bonfante & Genre, 2010). Two main groups of mycorrhizal symbioses include the arbuscular mycorrhizas of Glomeromycotina (Mucoromycota) and ectomycorrhizas, which are formed predominantly by members of Ascomycota and Basidiomycota. Ectomycorrhizal (EcM) lifestyle has evolved independently numerous times among fungi through convergent evolution (Martin *et al.*, 2016; Strullu-Derrien *et al.*, 2018), which involves a number of genomic innovations and metabolic modifications, including an increased genome size from expansions of transposable elements (TEs), a decrease in carbohydrate-active enzymes (CAZymes) involved in plant cell wall degradation, and diversification of small secreted proteins (SSPs) that may function as effectors in the mycorrhizal symbiosis (Kohler *et al.*, 2015).

One of the more enigmatic groups of EcM fungi is Endogonales. These fungi are members of Mucoromycota, one of the phyla formerly classified as zygomycetes (Spatafora *et al.*, 2016). Endogonales are unique, in that some members form small truffle-like sporocarps that contain zygospores (Supporting Information Fig. S1), one of the rare cases of sporocarp production with

Mucoromycota (Smith *et al.*, 2013). Endogonales comprise two families: Endogonaceae with *Endogone* and *Jimgerdemannia*, and Densosporaceae with *Densospora* and *Sphaerocreas* (Desirò *et al.*, 2017). In addition, both families include considerable phylogenetic diversity known only from environmental samples, as well as some arbuscule-forming fine root endophytes (Bidartondo *et al.*, 2011; Desirò *et al.*, 2017; Orchard *et al.*, 2017; Walker *et al.*, 2018).

Ecologically, Endogonales are characterized as EcM fungi of woody plants (Walker, 1985; Warcup, 1990; Yamamoto *et al.*, 2017), biotrophic symbionts of some nonspermatophytes (e.g. ferns, lycophytes, hornworts, and liverworts; Bidartondo *et al.*, 2011; Desirò *et al.*, 2013; Hirose *et al.*, 2014; Rimington *et al.*, 2015), and putative saprotrophs (nonsymbionts) associated with decayed wood, leaf litter, or rotten Polyporaceae basidiocarps (Berch & Fortin, 1983; Walker, 1985). Recently, the species group of *Glomus tenue*, the fine root endophytes of some vascular plants, were resolved as part of the Endogonales clade as well (Desirò *et al.*, 2017; Orchard *et al.*, 2017) and were placed into a new genus *Planticonsortium* (Walker *et al.*, 2018). This finding is based on short ribosomal DNA (rDNA) sequences alone, however, and will benefit from being tested by additional phylogenetic markers. EcM species form symbiotic structures consistent with mycorrhizal function, including a mantle of hyphae and a Hartig net. However, the function of Endogonales associated with woody plants remains undetermined (Tedersoo & Smith, 2017), as do the functions of fungal symbionts of hornworts, ferns, and lycophytes (van der Heijden *et al.*, 2015). Genome data will be essential for understanding the symbiotic functioning of these fungi.

It is important to mention that Mucoromycota also includes the arbuscular mycorrhizas of Glomeromycotina. In both Endogonales and Glomeromycotina there are some species known to harbor *Mycoplasma*-related endobacteria (MRE) in their hyphae and spores. Although these relationships are facultative for the fungi, the MRE are supported as phylogenetically specialized to Endogonales, Glomeromycotina, and Mortierellomycotina (Mucoromycota), and there has been increasing interest in understanding the distribution and evolution of these enigmatic bacteria (Bonfante & Desirò, 2017; Desirò *et al.*, 2018).

In addition to numerous independent origins, mycorrhizal symbioses are ancient. Plant fossils of the Rhynie Chert provide evidence of fungi associated with early land plants. The fossilized fungi share remarkable similarity with extant arbuscular mycorrhizas of Glomeromycotina, a finding that lends support for the hypothesis of Pirozynski & Malloch (1975) that colonization of terrestrial ecosystems by the green plant lineage may have been facilitated by the mycorrhizal symbiosis. The association of Endogonales with liverworts and hornworts and their placement in Mucoromycota have led to the hypothesis that they may be among the earliest forms of the mycorrhizal symbiosis, possibly as old as or older than Glomeromycotina (Field *et al.*, 2015a; Hoysted *et al.*, 2018). Rhynie Chert Glomeromycotina fossils, based on spores, vesicles, and arbuscules, date to *c.* 410 million yr ago (Ma) (Kenrick & Strullu-Derrien, 2014). The oldest definitive Endogonales fossils, which include diagnostic zygospores of Endogonaceae fungi, are from the Triassic and date to *c.* 240 Ma (Krings *et al.*, 2012). There are hyphal fossils from Rhynie Chert that have been interpreted as Mucoromycota, but the lack of fossilized spores associated with the material makes a definitive determination and lineage affiliation difficult (Strullu-Derrien *et al.*, 2014).

To address a number of hypotheses on the evolution of mycorrhizal symbioses and trophic ecology in Endogonales and Mucoromycota, we sequenced the genomes of four collections of Endogonaceae, including three EcM collections and one putative saprotroph. Because Endogonales are difficult to obtain and maintain in culture, these fungi were sequenced from fresh field-collected sporocarps; thus, the genome data generated are metagenomes of each fungus and its microbiome. This study had three main objectives. First, to use genome-scale phylogenetic analyses to confirm the placement of Endogonaceae within Mucoromycota and determine the ages of the most recent common ancestors (MRCAs) of Endogonaceae and Endogonales. These data can be used to compare the origin of Endogonales with that of Glomeromycotina to test whether Endogonales was a potential symbiont of the earliest land plants. Second, to determine whether Endogonaceae genomes show signatures of the EcM lifestyle, including large genome size due to TE expansions, reduction in the diversity of plant-cell-wall-degrading enzymes (PCWDEs), and diversification of SSPs that might function in the mycorrhizal symbiosis. EcM fungi within Ascomycota and Basidiomycota evolved from decomposers of plant substrates and show a striking decrease in PCWDEs, but the ecological context of Endogonaceae evolution has not been characterized in a phylogenomic paradigm. Third, to characterize the microbial communities associated with Endogonaceae sporocarps and determine whether metagenomes of Endogonaceae harbor MRE symbionts.

Materials and Methods

Genome sequencing, assembly, and annotation

Four Endogonaceae collections were selected for genome sequencing: *Endogone* sp. FLAS F-59071 was collected from a decayed log at the Ordway-Swisher Biological Station near Melrose, FL, USA (February 2018); *Jimgerdemannia flammicorona* AD002 was collected from soil under *Pinus strobus* at Veglio, Pemonte, Italy (September 2013); *J. flammicorona* GMNB39 was collected from soil under *Picea abies* and *Picea pungens* at Haslett, MI, USA (September 2015); and *Jimgerdemannia lactiflua* OSC 162217 was collected from a *c.* 40-yr old *Pseudotsuga menziesii* stand at Benton County, OR, USA (February 2015). For each collection, two to three sporocarps were used for DNA preparation. The sporocarps were cleaned and surface sterilized using 10% sodium hypochlorite and flash frozen with liquid N_2. Total DNA was extracted with a CTAB-based protocol following Mujic (2015). For *J. lactiflua* OSC 162217, a 150 bp pair-end library was constructed with New England Biomedical NEBNext kit and was sequenced with Illumina HiSeq3000. For each of *J. flammicorona* AD002, *J. flammicorona*

GMNB39, and *Endogone* sp. FLAS F-59071, a 150 bp pair-end library was constructed using the KAPA-Illumina library kit (KAPA Biosystems, Wilmington, MA, USA) and was sequenced with an Illumina HiSeq2500 sequencer.

Transcriptome sequencing was performed for *J. flammicorona* AD002, *J. flammicorona* GMNB39, and *Endogone* sp. FLAS F-59071. Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen). Messenger RNA was purified from total RNA using magnetic beads containing poly-T oligos. Standard complementary DNA libraries were generated from the purified messenger RNA using the Illumina Truseq Stranded RNA LT kit. Sequencing of the complementary DNA libraries was performed on the Illumina HiSeq2500 sequencer. The raw RNA-Seq reads were assembled using Trinity v.2.5.1 (Grabherr *et al.*, 2011).

Raw genomic Illumina reads were assembled using SPAdes v.3.8.1 (Bankevich *et al.*, 2012) with '-meta' option to generate assembly v.1. For each assembly, contigs longer than 1 kb were subjected to Blastn searches against the PATRIC database (Wattam *et al.*, 2014). Sequences of the 10 most frequently hit bacterial genomes were downloaded from the PATRIC database and were divided into 1 kb long fragments using a custom script. These fragmented bacterial genome sequences were used to spike the binning analysis of assembly v.1 using VizBin (Laczny *et al.*, 2015). We also spiked the binning analysis with fragmented sequences from four known MRE genomes from Glomeromycotina (Naito *et al.*, 2015; Torres-Cortés *et al.*, 2015) and one from Mortierellomycotina (A. Desirò and G. Bonito, personal communication) to identify potential MRE sequences in assembly v.1. Sequences clustered with the non-MRE bacterial marker sequences were removed from assembly v.1. Sequences clustered with the known MRE sequences were removed from assembly v.1 and processed as MRE sequences in the subsequent analysis. Additional contigs were removed if they formed isolated clusters from the main body of data and showed strong affinities to non-fungal sequences in the subsequent Blastx searches against the National Center for Biotechnology Information (NCBI) nr database. After removal of nontarget sequences, the remaining assembly v.1 contigs were saved as assembly v.2.

Original raw reads were aligned to assembly v.2 using Bowtie2 v.2.2.3 (Langmead & Salzberg, 2012). The mapped reads were reassembled with SPAdes v.3.8.1 (Bankevich *et al.*, 2012) to generate assembly v.3. A VizBin (Laczny *et al.*, 2015) analysis on assembly v.3 divided the contigs into two regions: gene space and a TE-rich region, with the former containing the majority of contigs with non-TE protein-coding genes and the latter composed of mainly TE-related sequences (Fig. S2). Contigs from the gene space region of assembly v.3 were used to query the UniProt-TrEMBL database (Apweiler *et al.*, 2004) and a database of 347 fungal proteomes using Diamond v.0.8.38 (Buchfink *et al.*, 2014) with the Blastx option. Nonfungal contigs and non-Mucoromycotina fungi contigs were removed and the remaining contigs were considered the final working assembly (Fig. S3). The completeness of the final genome assemblies was assessed using Busco (Simão *et al.*, 2015).

Genome annotation was performed using the Maker v.2.10 annotation pipeline (Cantarel *et al.*, 2008; Holt & Yandell, 2011). To facilitate protein model prediction and annotation, we included a dataset of predicted protein models from the seven fungal species *Chytridium lagenaria*, *Mortierella multidivaricata*, *Neurospora crassa*, *Rhizophagus irregularis*, *Rhizopus oryzae*, *Syncephalis plumigaleata* and *Umbelopsis ramanniana* as protein homology evidence. Transcriptome data were used to facilitate annotation of *J. flammicorona* AD002, *J. flammicorona* GMNB39 and *Endogone* sp. FLAS F-59071. For the annotation of *J. lactiflua* OSC 162217, which lacked transcriptome data, we used the transcriptome data from the other three Endogonales collections as the alternative expressed sequence tag data in the Maker pipeline.

Phylogenetics of Endogonaceae and molecular dating analysis

We sampled a total of 66 species (63 fungi and three outgroup taxa), including 55 assembled and annotated genomes and 11 transcriptomes (Table S1). We employed 434 protein markers that have previously proved useful for higher level phylogenetic analysis of fungi (Beaudet *et al.*, 2018; https://github.com/1KFG/Phylogenomics_HMMs). The search for the target markers in the sampled genomes and proteomes and the alignments was performed using the pipeline PHYling (Stajich JE; http://github.com/stajichlab/PHYling_unified) with the default setting. Phylogenetic analyses were performed on the concatenated alignment of the 434 markers in RAxML v.8.0.26 (Stamatakis, 2014). We employed the '-f a' option with 100 bootstrap replicates and the PROTGAMMALG model. To facilitate subsequent analysis on CAZymes and SSPs, we removed 11 taxa that had only transcriptome data and one taxon with no annotation data from the dataset (Table S1). We then performed RAxML analysis with the remaining 54 taxa using the same parameters just described.

To date the divergence times in the best RAxML tree, we used r8s v.1.8 (Sanderson, 2003) with a local molecular clock and a total of eight possible calibration points. The MRCA of all fungi except for Blastocladiomycota and *Rozella* was constrained to 750 Ma based on the expansion of fungal pectinases and the age of streptophytes as discussed in Chang *et al.* (2015). An additional seven fossils were used as minimum age calibration points for the MRCAs of Blastocladiomycota (407 Ma), Chytridiomycota (407 Ma), Glomeromycotina (407 Ma), Endogonaceae (247 Ma), Mucorales (315 Ma), Ascomycota (407 Ma), and Basidiomycota (330 Ma) (Table S2). Fossil calibration points were cross-validated by fixing the MRCA of Ascomycota along with the MRCA of Chytridiomycota and all other fungi. After passing cross-validation, the MRCA of Ascomycota was reset to minimum age with final settings as follows: divtime method = PL, algorithm = TN, set smoothing = 10, penalty = log, minRateFactor = 0.5, checkgradient = yes.

A second dating analysis was performed based on a four-locus dataset modified from Desirò *et al.* (2017, Fig. 1). This dataset has a wider taxon sampling in Endogonales, including four Densosporaceae and 52 Endogonaceae collections. The four markers applied were 18s rDNA, 28s rDNA, translation elongation factor EF1A, and ribosomal polymerase II core subunit RPB2. The

J. lactiflua OSC 162217 collection was not included in this analysis due to the absence or poor quality of the sequences of the target markers. We included *Mortierella verticillata* NRRL 6337 as the outgroup. Dating analysis was performed using BEAST v.2.4.8 (Bouckaert *et al.*, 2014) with a random starting tree. The minimum age of the MRCA of Endogonaceae was set as 247 Ma.

Identification and characterization of TEs

We used REPEATSCOUT v.1.0.5 (Price *et al.*, 2005) and LTRHARVEST (Ellinghaus *et al.*, 2008) to carry out *de novo* identification of repetitive sequences in the four Endogonaceae genomes. We then applied multiple filters to the output files of REPEATSCOUT and LTRHARVEST to remove false positives. For the repeat library derived from REPEATSCOUT, we filtered out low-complexity and tandem repeats that occurred < 10 times in the genome or had a sequence length of < 100 bp. For the LTRHARVEST output files, we removed repeats shorter than 400 bp, those with fewer than five BLASTN hits to the genome (significance cutoff of 1e-15), and those with BLASTX hits to genes in UniProt/SwissProt (significance cutoff of 1e-5). The filtered output files from REPEATSCOUT and LTRHARVEST were consolidated and deduplicated at the 80% similarity level using USEARCH v.4.2.66 (Edgar, 2010). The final set of repeat elements were annotated by TBLASTX searches (significance cutoff of 1e-15) against REPBASE 21.06 (Bao *et al.*, 2015) and classified into major TE families with custom scripts.

Taxonomic profiling of Endogonaceae microbiomes

Taxonomic profiling of the microbiome associated with each assembly was performed using METAPHLAN2 (Truong *et al.*, 2015). To remove reads identified as Endogonaceae, raw reads were aligned to the final assemblies using BOWTIE2 v.2.2.3 (Langmead & Salzberg, 2012). The unmapped reads were extracted and aligned to MRE contigs identified in the VIZBIN analyses. The reads mapped to MRE contigs were removed and the unmapped reads were analyzed using METAPHLAN2 with default settings to investigate the composition and relative abundance of bacteria at class and order levels.

Phylogenetics of MRE

Using previously published MRE 16S rDNA sequences as queries, we performed BLASTN searches against the MRE sequences identified in this study with a significance cutoff of 1e-80. The newly retrieved 16S rDNA sequences were aligned using MAFFT (Katoh & Standley, 2013), together with 58 published 16S rDNA sequences from other Mucoromycota-associated MRE, eight from Tenericutes bacteria, and two from Firmicutes bacteria. The resulting alignment was then trimmed with GBLOCKS v.0.91b (Castresana, 2000) using the least stringent conditions. Phylogenetic reconstructions were carried out with MRBAYES v.3.2.6 (Ronquist *et al.*, 2012) and RAXML v.8.2.4 (Stamatakis, 2014). The best-fit nucleotide substitution model was estimated with JMODELTEST v.2.1.9 (Darriba *et al.*, 2012). Markov chain Monte Carlo was run for 10 million generations under the TrN+I+G model, whereas maximum likelihood analysis was conducted with the autoMR option of automatic 'bootstrapping' (Pattengale *et al.*, 2010) under the GTRCAT model.

Comparative genomic analysis on CAZymes and SSPs

For the identification and characterization of CAZymes and SSPs, we excluded 11 sampled taxa with only transcriptome data and one with only genome but no predicted proteome data (Table S1). We focused on 45 families and subfamilies of PCWDEs for the CAZyme analyses. The dbCAN enzyme hidden Markov model (HMM) profiles (http://csbl.bmb.uga.edu/dbCAN/; Yin *et al.*, 2012) were used to query the 54 predicted proteomes using *hmmscan* in HMMER v.3.1b (Eddy, 2011). The *hmmscan* results were filtered with an *e*-value cutoff of 1e-15 and an alignment coverage cutoff of 50% of the domain query. The sequences for each of the enzyme families and subfamilies were aligned using *hmmalign* and trimmed with TRIMAL (Capella-Gutiérrez *et al.*, 2009). Individual gene trees were reconstructed using RAxML 8.0.26 (Stamatakis, 2014) with '*-fa*' option, 100 bootstrap replicates, and a protein model of PROTGAMMALG. Species-gene tree reconciliation was performed with NOTUNG v.2.9 (Durand *et al.*, 2005; Stolzer *et al.*, 2012), using the best RAxML tree based on 54 taxa and 434 markers as the species tree. To accommodate the uncertainties in gene tree phylogenies, branches in the gene trees with < 70% bootstrap support were collapsed and rearranged to obtain the most parsimonious reconciliation with the *-rearrange* option. We also screened for all the 360 dbCAN families with *hmmscan* and filtered the results with the same setting as already described. We performed a two-way clustering based on the distribution of enzyme copy number across the sampled 54 taxa using the HEATMAP.2 program in the R package. We also investigated the distribution of class II peroxidases (PODs) in the sampled taxa. We downloaded 267 fungal POD sequences from the PeroxiBase (http://peroxibase.toulouse.inra.fr) and built HMM profiles using *hmmbuild*. We searched for POD sequences from the 54 predicted proteomes using *hmmscan* and filtered the results with an *e*-value cutoff of 1e-15 and an alignment coverage cutoff of 50% of the query. The resulting sequences were searched against the UniProt-TreEMBL database using DIAMOND *blastp*. Only the sequences with the best hit to a class II peroxidase were considered as POD sequences.

The identification of the secretome and SSPs was performed utilizing previously developed pipelines (Pellegrin *et al.*, 2015; Lu, 2016). The following filters were applied to the predicted proteomes of each Endogonaceae genome: (1) possession of a signal peptide as predicted by SIGNALP v.4.1 with sensitivity cutoff of 0.34; (2) targeted for a secretory pathway as predicted by TARGETP (Emanuelsson *et al.*, 2000); (3) absence of a transmembrane helix as predicted by TMHMM v.2.0 (Krogh *et al.*, 2001); (4) targeted for extracellular localization as predicted by WOLF PSORT v.0.1 (Horton *et al.*, 2007); (5) not targeted to the endoplasmic reticulum as predicted by PS-SCAN (http://www.hpa-bioinfotools.org.uk/cgi-bin/ps_scan/ps_scanCGI.pl) under the Prosite accession of PS00014 (PDOC00014). We considered a secretome protein as an SSP if it was shorter than 300 amino acids.

Results

Genome assembly and data binning

The oligonucleotide-composition-based binning process removed 11–34% of the data from the initial assemblies, including sequences from MRE, other bacteria, and various eukaryotes. The genome assembly sizes vary from 96 Mb for *Endogone* sp. FLAS F-59071, 180 Mb for *Jimgerdemannia lactiflua* to *c.* 230 Mb for the two *J. flammicorona* collections. All four genomes have a high proportion of repetitive elements, and the GC contents of the repetitive elements are lower than those of the nonrepeat regions (Table 1). Our assemblies captured 76.3–84.15% of the 290 fungal universal single-copy orthologues included in the BUSCO database (Simão *et al.*, 2015). Among the four genomes, *Endogone* sp. FLAS F-59071 has the least number of predicted gene models (9569; Table 1), followed by *J. lactiflua* OSC 162217 (12 651). The two *J. flammicorona* collections have slightly more predicted gene models (13 838 and 13 653).

Repetitive elements

The identified repetitive elements of the Endogonaceae genomes comprised 65.6–78.1% of the total assemblies. Class I transposons (retrotransposons) represent the largest groups of repetitive elements in all four genomes (33.3–54.9% of all the repeats; Table S3). Within class I transposons, Gypsy long terminal repeat (LTR) was the most abundant group, accounting for nearly or over half of all class I transposons, followed by Copia LTR (Table S3). Class II transposons comprised a small proportion of the total repeats (1.2–6.4%). A large proportion of the *Endogone* sp. FLAS F-59071 genome was made up of simple repeats (31.3%), whereas in the other genomes the proportions of simple repeats were much lower (12.0–15.9%).

Microbiome and *Mycoplasma*-related endobacteria

The binning analysis using VIZBIN (Laczny *et al.*, 2015) identified MRE sequences from three of the four Endogonaceae genomes (Fig. 1), with the exception being *J. flammicorona* GMNB39 for which no MRE sequences were detected. The total numbers of nucleotides in the MRE contigs were 0.9 Mb for *J. flammicorona* AD002, 1.8 Mb for *J. lactiflua* OSC 162217, and 2.9 Mb for *Endogone* sp. FLAS F-59071 (Table 2). This is proportional to the number of MRE phylotypes identified from these genomes using 16S rDNA sequences, with one MRE phylotype from *J. flammicorona* AD002, two from *J. lactiflua* OSC 162217, and three from *Endogone* sp. FLAS F-59071 (Table 2). Phylogenetic reconstructions based on 16S rRNA gene sequences placed two MRE phylotypes from *Endogone* sp. FLAS F-59071 and one from *J. lactiflua* OSC 162217 within a clade encompassing MRE sequences previously retrieved from other Endogonaceae sporocarps (Desirò *et al.*, 2015). The remaining three MRE phylotypes, one from each of the three Endogonaceae collections, formed a novel early-diverging clade (Fig. 2).

After the removal of reads from MRE, Proteobacteria dominated the microbiomes of both *J. flammicorona* collections (96.7-% and 99.8%), although the bacterial composition at order level differed between them (Fig. S4). In addition to Proteobacteria, the *J. lactiflua* OSC 162217 microbiome contained 44% Actinobacteria, whereas the *Endogone* sp. FLAS F-59071 microbiome contained 26.9% Acidobacteria. The microbiome associated with *Endogone* sp. FLAS F-59071 was the most taxonomically diverse, with seven bacterial orders and one DNA viral order (Fig. S4).

Phylogenetic placement and dating of Endogonaceae and Endogonales

The RAxML (Stamatakis, 2014) analysis was based on 66 taxa and 434 conserved protein regions, which incorporated 128 068 aligned amino acid sites. The best maximum likelihood tree received strong bootstrap support across the phylogeny, with a few exceptions (Fig. 3). Reconstructed relationships among and within major fungal lineages from our analysis are consistent with previous studies (Spatafora *et al.*, 2016; Torres-Cruz *et al.*, 2017). Endogonaceae is monophyletic and placed within Mucoromycotina as sister group to the clade of Mucorales, Umbelopsidales, and *Bifiguratus adelaidae*. Within Endogonaceae, the two *J. flammicorona* collections are closest relatives to each other, forming a clade sister to *J. lactiflua* OSC 162217. *Endogone* sp. FLAS F-59071 was resolved as sister to *Jimgerdemannia*.

The dating analysis with R8S (Sanderson, 2003; Taylor & Berbee, 2006) estimated that the age of the MRCA of Mucoromycota as 620 Ma, the age of the MRCA of Glomeromycotina as

Table 1 Statistics of genome assemblies and annotations for *Endogone* sp. FLAS F-59071, *Jimgerdemannia flammicorona* AD002, *J. flammicorona* GMNB39, and *Jimgerdemannia lactiflua* OSC 612217.

	Endogone sp. FLAS F-59071	*J. flammicorona* AD002	*J. flammicorona* GMNB39	*J. lactiflua* OSC 612217
Size (Mb)	96	231	240	180
Total no. of contigs	15 409	35 354	33 875	45 525
No. of contigs with gene models	3861	6385	6313	7875
No. of predicted gene models	9569	13 838	13 653	12 651
BUSCO (%, complete/partial models)	84.1 (76.9/7.2)	83.8 (70.0/13.8)	81.4 (66.9/14.5)	76.3 (56.6/19.7)
Masked repeats (% total assembly)	65.6	77.6	78.1	72.9
Overall GC%	39.0	42.8	42.7	44.4
Non_repeat GC%	44.9	45.2	45.0	46.8
Repeat GC%	35.4	41.4	41.5	43.0

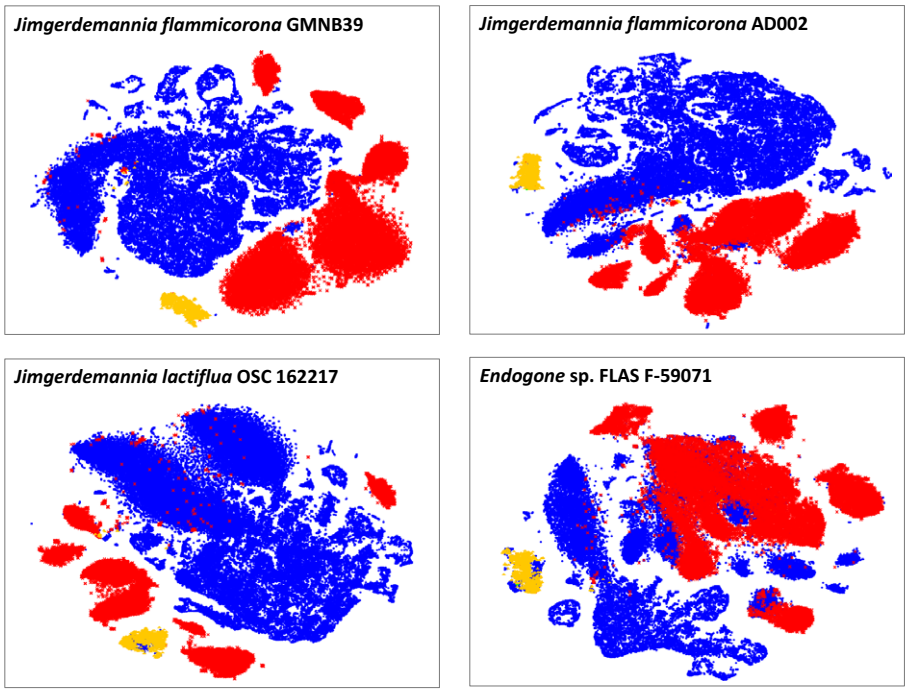


Fig. 1 Metagenomic data binning of the four Endogonales genomes based on tetramer nucleotide composition using VizBin. Contigs longer than 1 kb in assembly v.1 for each fungal isolate were included in the analyses (blue data points). The analyses were spiked with known *Mycoplasma*-related endobacteria sequences (yellow data points) and sequences from the bacterial genomes that showed up most frequently (top 10) in the Blastn search of assembly v.1 contigs against the PATRIC database (red data points). Sequences from assembly v.1 that clustered with spiked bacterial sequences were removed from the assembly as contaminants. Additional isolated islands of contigs of assembly v.1 were sampled and used in Blastp analysis against the NCBI nr database. Those showing nonfungal affinity were removed from the assembly.

426 Ma, and the age of the MRCA of Endogonaceae as 258 Ma (Fig. 3). Beast (Bouckaert *et al.*, 2014) analysis on the four-locus dataset estimated the age of MRCA of Endogonaceae as 251 Ma, the age of MRCA of Densosporaceae as 167 Ma, and the age of MRCA of Endogonales as 420 Ma (Fig. 4).

Comparative genomic analysis

The 45 enzyme families and subfamilies of PCWDEs were divided into five enzyme classes, including auxiliary activities (AAs, with only AA9 included in this study), carbohydrate-binding modules (CBMs, with only CBM1 included in this study), carbohydrate esterase, glycoside hydrolases (GHs), and polysaccharide lyases. In addition, we compared the lignin-degrading POD repertoire in our study. Overall, Mucoromycota species possess fewer PCWDEs than Dikarya fungi do (Fig. 3; Table S4). Enzyme families targeting crystalline celluloses (AA9, GH6, and GH7) and lignin-degrading PODs were absent from Mucoromycota taxa (Fig. 3; Table S4). CBM1 modules, which are also involved in the degradation of cellulose, were absent or present with low copy numbers in Mucoromycota species. There was one copy of CBM1 present in the *J. flammicorona* AD002 genome, whereas none were identified from the other three Endogonaceae genomes. The MRCA of Mucoromycota was inferred to have 48 PCWDEs (Fig. 3; node 67 in Fig. S5 and in Table S4). The number of PCWDEs remained stable, with a small number of duplication and loss events along the backbone of Mucoromycota (Fig. 3; Tables S5, S6). By contrast, we reconstructed 23 loss events and no duplications on the branch leading to the MRCA of Endogonaceae. As a result, the MRCA of Endogonaceae is predicted to possess only 23 copies of PCWDEs. Despite the small number, PCWDEs from each of the sequenced Endogonaceae genomes comprise enzymes targeting a wide range of polymers found in plant primary cell walls, including cellulose, pectin, xylan, and mannose (e.g. GH5–7, GH9, GH10, and GH28).

The Endogonaceae secretomes account for 3.9–4.8% of the corresponding proteomes (Fig. 5; Table S7), comparable to the

Table 2 Statistics of the *Mycoplasma*-related endobacteria (MRE) sequences identified from the four sequenced Endogonales metagenomes.

Fungal isolate name	Total no. of MRE contigs	Total no. of MRE nucleotides (Mb)	Total no. of MRE phylotypes
Endogone sp. FLAS F-59071	532	2.9	3
Jimgerdemannia flammicorona AD002	64	0.9	1
J. flammicorona GMNB39	0	0	0
Jimgerdemannia lactiflua OSC 612217	279	1.8	2

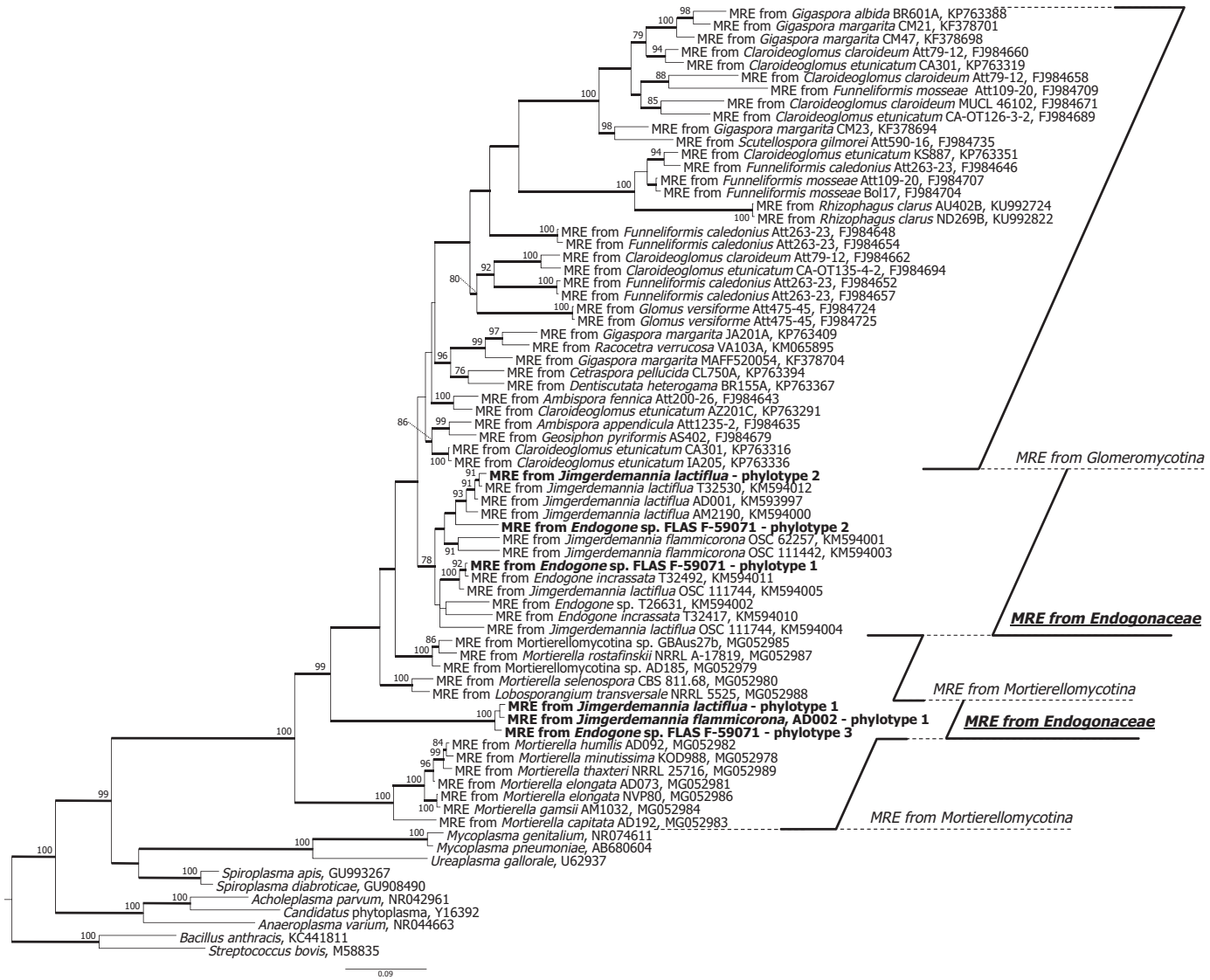


Fig. 2 Phylogenetic placement of *Mycoplasma*-related endobacteria (MRE) phylotypes identified in the Endogonales specimens of *Endogone* sp. FLAS F-59071, *Jimgerdemannia flammicorona* AD002, and *J. lactiflua* OSC 162217 based on MrBayes analysis of 16S rDNA sequences. MRE phylotypes detected in this study were found in two Endogonales-specific clades. The first clade encompassed three MRE phylotypes identified in this study as well as MRE sequences previously reported from other Endogonaceae specimens. The remaining three MRE phylotypes identified in this study, one phylotype from each of three Endogonales specimens, formed an early-diverging clade among all the sampled MRE sequences. The tree branches with Bayesian posterior probabilities ≥ 0.95 are thickened and maximum likelihood bootstrap support values ≥ 70 are shown. MRE sequences obtained in this study are in bold.

average secretome/proteome ratios reported for other fungi by Kim *et al.* (2016). The proportions of SSPs in the secretome for the Endogonaceae genomes varied from 48% to 57%. The SSP/secretome ratios of Endogonaceae are similar to average values for the sampled Mucoromycota species (51%). SSP/secretome ratios for the two EcM species from Ascomycota (*Tuber melanosporum* and *Cenococcum geophilum*) are comparable to those for Endogonaceae (Fig. 5; Table S7). On the other hand, the four EcM species from Basidiomycota (*Amanita muscaria*, *Laccaria bicolor*, *Pisolithus tinctorius*, and *Suillus brevipes*) possess higher SSP/secretome ratios (Fig. 5; Table S7), consistent with previously reported values (Pellegrin *et al.*, 2015; Kim *et al.*, 2016).

Discussion

Assembly and annotation of fungal genomes using a metagenome approach

The rapid sequencing of fungal genomes in recent years has expanded our knowledge of the diversity of the fungal tree of life and deepened our understanding of many aspects of fungal biology. Much of the sequencing effort has been focused on fungal species that can be cultured, but a large number of biotrophic fungal species cannot be maintained in culture. For these taxa, sporocarp tissues are often the only available material for generating genome data. Sporocarps are not

New Phytologist

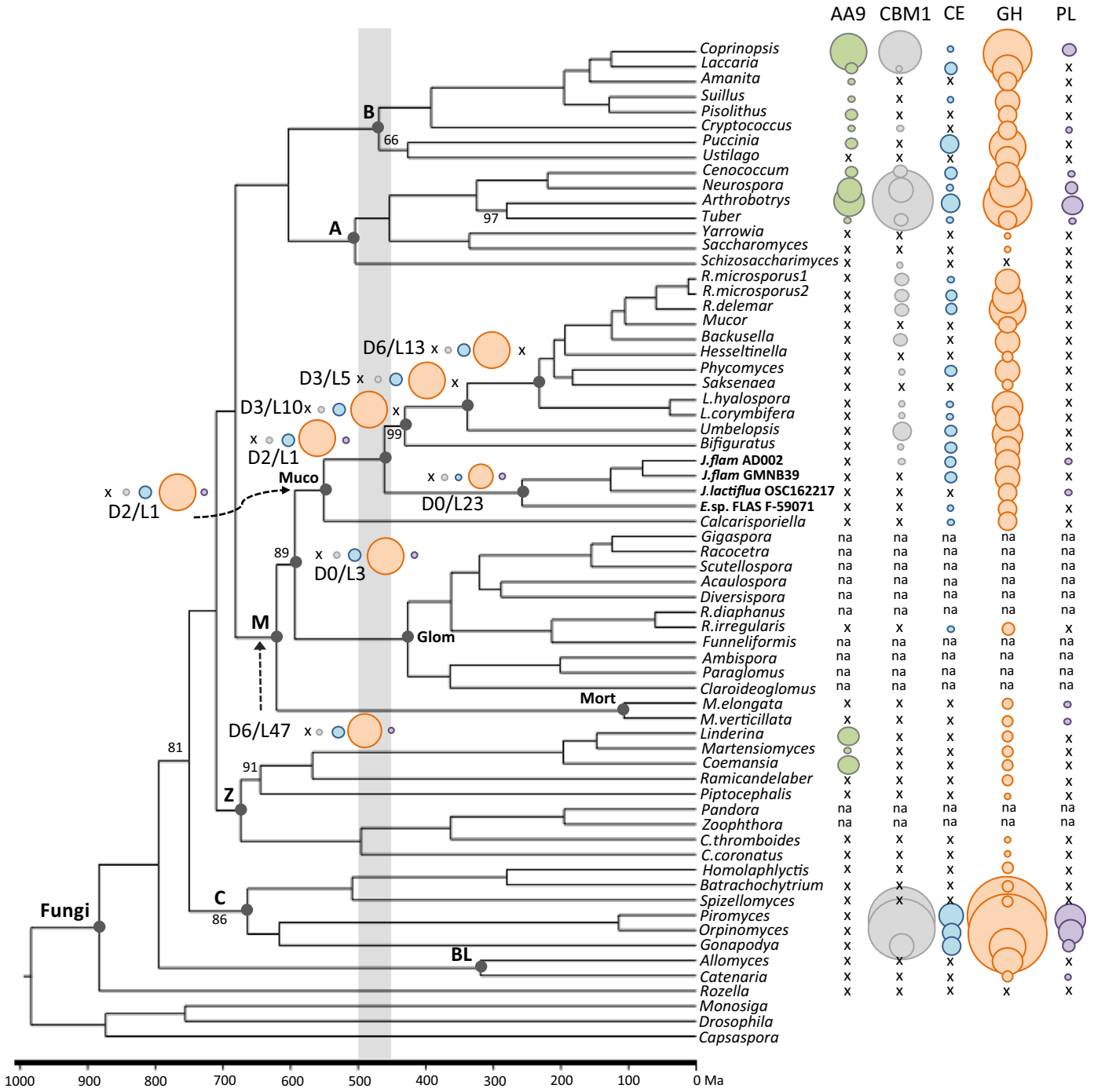


Fig. 3 Phylogenetic placement and divergence time estimation of Endogonales, and the evolution of fungal plant-cell-wall-degrading enzymes (PCWDEs). The four genomes sequenced for this study are highlighted in bold text. The chronogram is based on the best maximum likelihood tree inferred with RAxML (66 taxa, 434 proteins, and 128 068 aligned amino acid sites) and the divergence times are estimated with r8s. The vertical gray shade indicates the estimated origin of embryophytes (e.g. land plants). Bootstrap support lower than 100 is shown by the branches. Reconstruction of PCWDE evolution was performed with Notung. Colored circles on the right indicate the observed and reconstructed number of major categories of PCWDEs (AA9, auxiliary activity family 9; CBM1, carbohydrate-binding module 1; CE, carbohydrate esterases; GH, glycoside hydrolases; PL, polysaccharide lyases). The area of the circles is proportional to the number of PCWDE modules, with 'x' indicating zero and 'na' referring to the 12 taxa not included in Notung analysis because they only possess transcriptome data (Glomeromycotina and *Pandora*) or lack annotated proteome data (*Zoophthora*). Reconstructed numbers of PCWDE modules are shown for a selection of internodes of Mucoromycota, together with the numbers of reconstructed gene duplication and loss events (duplication/loss, D/L) for those nodes. Reconstructed numbers of PCWDE modules for other internodes are shown in Supporting Information Table S4. Node names: A, Ascomycota; B, Basidiomycota; BL, Blastocladiomycota; C, Chytridiomycota; M, Mucoromycota; Z, Zoopagomycota; Muco, Mucoromycotina; Glom, Glomeromycotina; Mort, Mortierellomycotina. The time scale is in millions of years ago (Ma).

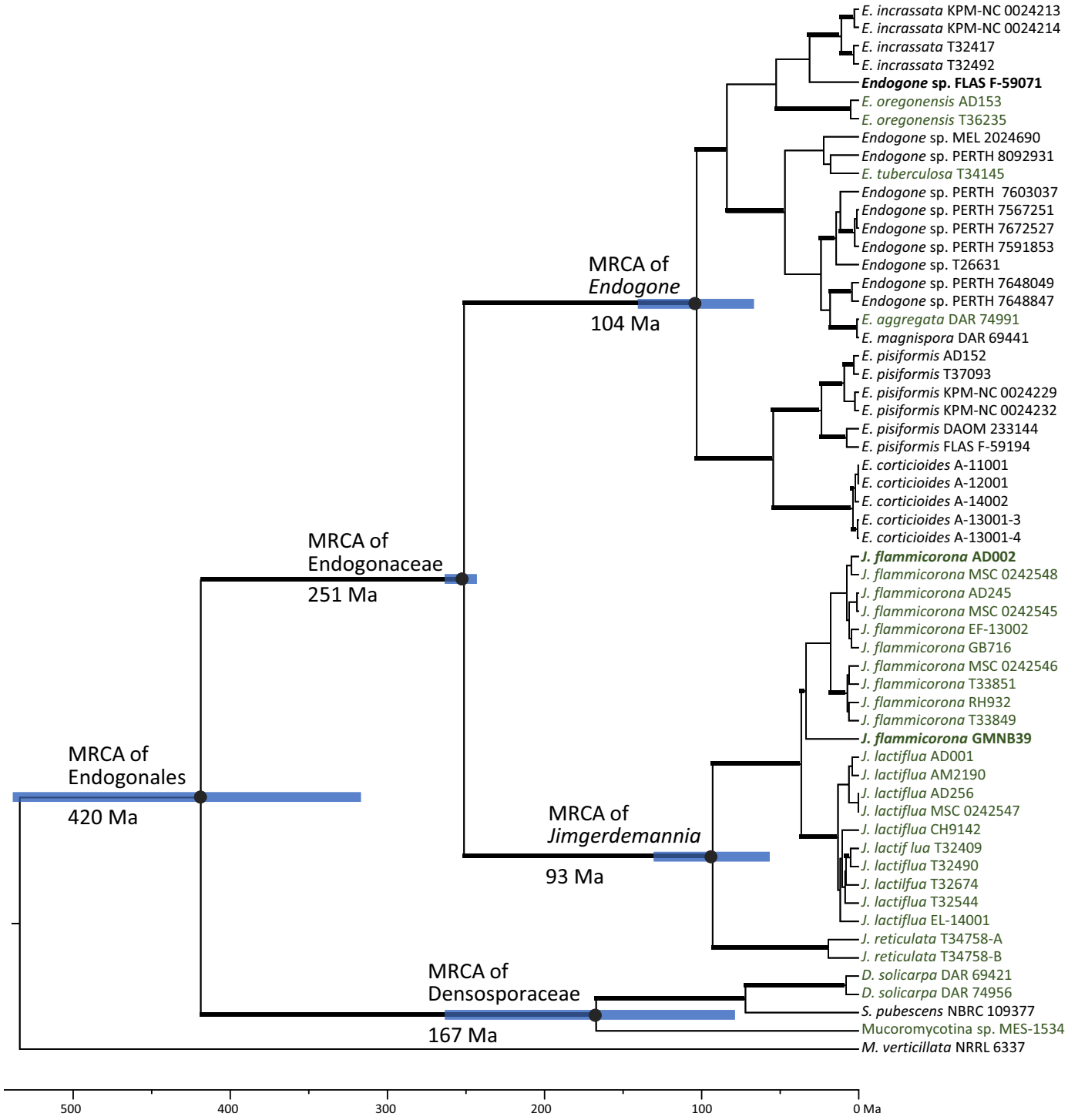
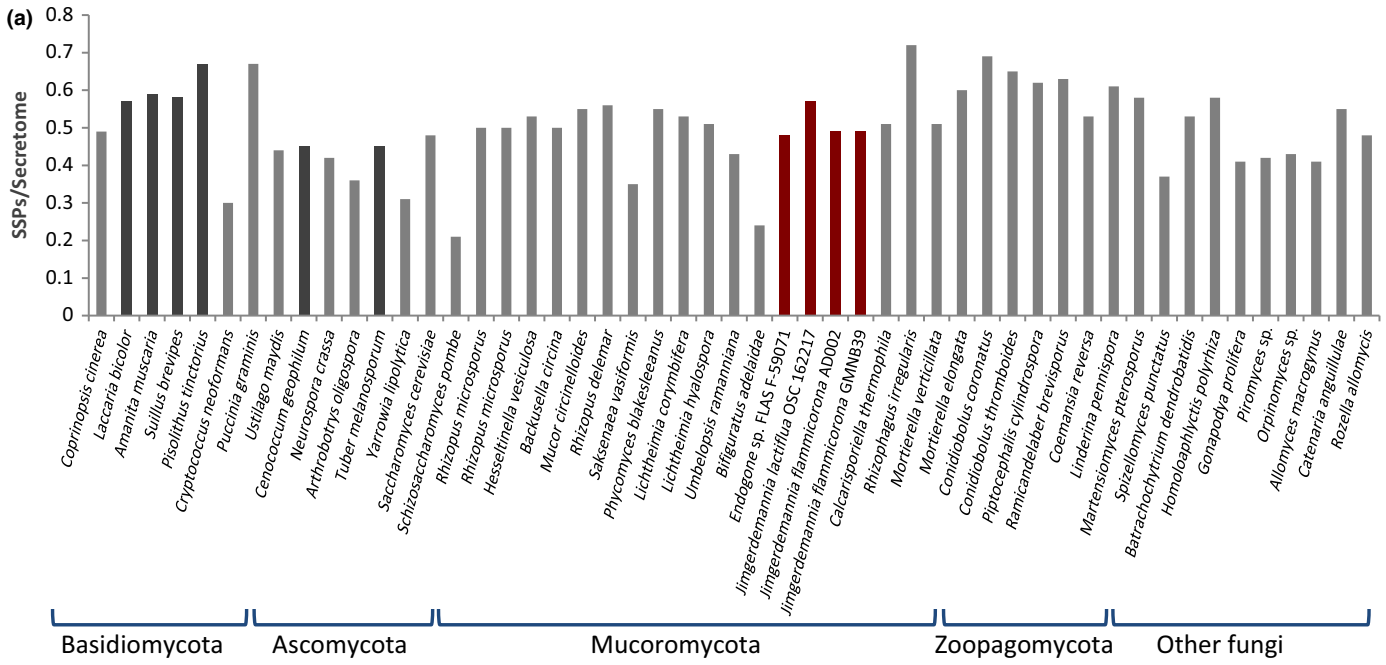
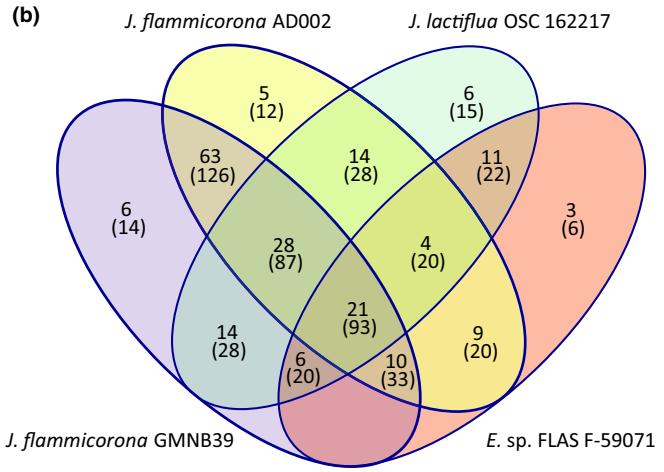


Fig. 4 Molecular dating analysis using BEAST v.2.4.8 on the origin and diversification of Endogonales based on the four-locus dataset (18s rDNA, 28s rDNA, EF1a and RPB2; 57 isolates from Endogonales and *Mortierella verticillata*). The genomes sequenced in this study are highlighted in bold text. No sequences from the *Jimgerdemannia lactiflua* OSC162217 genome are included due to the absence or poor quality of the target markers. Numbers near the nodes represent median node ages, and blue bars represent 95% highest posterior density around the nodes. Thickened branches denote those receiving bootstrap support higher than 70 in the RAxML analysis based on the same dataset. Taxon names in green denote the taxa with ectomycorrhizal (EcM) or putative EcM ecology. MRCA, most recent common ancestor. The time scale is in millions of years ago (Ma).

axenic, and sequencing from these structures will capture the metagenome of the sample. Bacteria and eukaryotes associated with fungal sporocarps may exist as symbionts, predictable members of the sporocarp microbiome, or in a more transient way as environmental contaminants (Quandt *et al.*, 2015). Careful bioinformatic mining of metagenomic data remains a major challenge in gaining access to a suitable fungal genome from sporocarp tissue.





(c)

	Total SSPs	SSPs as singletons	SSPs in species-specific clusters	Total SSSPs
J. flammicorona GMNB39	299	131	14	145
J. flammicorona AD002	316	147	12	159
J. lactiflua OSC 162217	280	162	15	167
E. sp. FLAS F-59071	222	146	6	152

Fig. 5 Comparative analysis of small secreted proteins (SSPs). (a) Proportion of SSPs in secretome across the sampled 54 predicted proteomes. SSP/secretome ratios are in red for the four Endogonales isolates, in black for the six Dikarya ectomycorrhizal species, and in gray for all the other taxa. Fungal taxa are arranged according to their phylogenetic placement. (b) Venn diagram of the SSP cluster distribution among the four Endogonales genomes. SSP clusters were identified using FastOrtho. The number in each division is the number of clusters in the division, and the numbers in parentheses are the total number of genes included in the clusters in the division. (c) Statistics of species-specific SSPs (SSSPs) in the four Endogonales genomes based on the results of FastOrtho analysis of SSPs.

In this study, we employed a binning approach based on oligonucleotide composition, guided and complemented with information from sequence coverage and Blast analyses. Oligonucleotide-composition-based binning methods have been widely used in microbial metagenome analyses (Sedlar *et al.*, 2017) but have only rarely been used for metagenomic data from eukaryotic genomes (e.g. Mosier *et al.*, 2016; López-Escardó *et al.*, 2017; Quandt *et al.*, 2017). In our analysis, we visually differentiated and identified sequences from nontarget organisms based on their unique oligonucleotide signatures (Fig. 1). We also successfully retrieved genome sequences for the associated MRE. The four Endogonaceae assemblies possessed 76–84% of the 290 fungal universal orthologues in the BUSCO database and 82–87% of the 434 markers used in the phylogenetic inference. Even though the assemblies comprised a large number of contigs (Table 1), gene models were present for only 17–25% of the contigs (Table 1). This approach captured gene space in a manner comparable to single cell genomics (Quandt *et al.*, 2017; Beaudet *et al.*, 2018) and was sufficient to address our original hypothesis. The core gene space and repeat-rich regions of Endogonaceae

genomes were also separated from each other in oligonucleotide composition space (Fig. S2), with repeat-rich contigs identified as more AT-rich (discussed later).

Metagenome binning and the microbiomes of Endogonaceae sporocarps

We identified MRE sequences from three of the four Endogonaceae genomes, a finding consistent with the facultative nature of the fungal symbiont. The total number of nucleotides identified as MRE is proportional to the number of MRE 16S phylotypes (Table 2). We estimate that the genome size of MRE to be *c.* 0.9 Mb, within the range of typical MRE genome sizes (Araldi-Brondolo *et al.*, 2017; Naito *et al.*, 2017). In line with Desirò *et al.* (2018), phylogenetic reconstructions based on 16S rDNA sequences identified several host-specific MRE clades, including one from Glomeromycotina, two from Mucoromycotina (Endogonaceae), and three from Mortierellomycotina (Fig. 2). Three of the MRE phylotypes identified in this study were resolved in the Endogonaceae-specific clade, whereas the other three MRE phylotypes, one from each Endogonaceae collection, formed a clade distantly related to an Endogonaceae-specific clade. They represent one of the earliest splits in the MRE phylogeny, together with the three Mortierellomycotina-specific clades (Fig. 2). The existence of this novel MRE clade suggests that additional MRE diversity is waiting to be discovered and that the origin and evolution of the fungus–MRE association might have been more complicated than previously hypothesized by Bonfante & Desirò (2017).

In contrast to vertically inherited endohyphal bacteria, extrahyphal bacteria are often acquired horizontally through the soil bacterial community (Mondo *et al.*, 2012). Not surprisingly, after excluding MRE, the most abundant bacterial phyla identified from each Endogonaceae collection were bacteria that are prevalent in rhizosphere soils (Martin *et al.*, 2017). The major bacterial phyla found in Endogonaceae are also the same taxa identified in microbiomes of the distantly related truffle genera *Tuber* and *Elaphomyces* (Antony-Babu *et al.*, 2014; Quandt *et al.*, 2015; Benucci & Bonito, 2016). By contrast, at ordinal level, the bacterial communities associated with Endogonaceae collections were unique among themselves and different from those found in *Tuber* and *Elaphomyces*, likely reflecting the differences in the local bacterial communities where the sporocarps were collected.

Phylogenetic relationships and dating analysis in Endogonales

Phylogenetic inferences at the phylum and subphylum levels in this study are consistent with recent genome-scale phylogenies (Spatafora *et al.*, 2016; Torres-Cruz *et al.*, 2017). Within Mucoromycotina, Endogonaceae is resolved sister to the *Bifiguratus*–Mucorales–Umbelopsidales clade (Fig. 5). The monophyly of Endogonaceae is well supported (Fig. 5), and the species relationships are consistent with previous studies based on fewer loci but more extensive taxon sampling (Desirò *et al.*, 2017). Both *J. flammicorona* and *J. lactiflua* are EcM species, whereas *Endogone* sp. FLAS F-59071 is a member of the *Endogone pisiformis* clade (Desirò *et al.*, 2017), previously reported as saprotrophic (Berch & Fortin, 1983). The origin of Endogonaceae was dated at the Permian–Triassic boundary, whereas the diversification of the extant members did not occur until after the Lower Cretaceous. The age of the MRCA of Endogonales was estimated as mid–late Silurian and is similar to the estimated MRCA age of Glomeromycotina (Figs 3, 4).

It had been hypothesized that the initial colonization of land by plants was facilitated through the symbiotic interactions between plants and Glomeromycotina fungi (e.g. Pirozynski & Malloch, 1975; Selosse & Le Tacon, 1998). This paradigm was recently challenged with the discovery of mycorrhizal-like associations between Mucoromycotina fungi and liverworts and hornworts (Bidartondo *et al.*, 2011; Desirò *et al.*, 2013; Field *et al.*, 2015b). This discovery has led to the hypothesis that Mucoromycotina fungi may also have played an important role in the colonization of land by plants (Field *et al.*, 2015a). Mucoromycotina fungi that are symbiotic with nonspermatophyte plants were phylogenetically placed within Endogonales, in both Endogonaceae and Densosporaceae (Bidartondo *et al.*, 2011; Rimington *et al.*, 2015; Desirò *et al.*, 2017). We estimate that Endogonales originated in the mid–late Silurian (~420 Ma), contemporary to the origin of Glomeromycotina. This suggests that the establishment of the Endogonales–plant association may have co-occurred around the same time as the arbuscular mycorrhizal association, both during the initial colonization of land by plants. Unfortunately, our sampling does not include any species of Endogonales that form mycorrhizal-like associations with nonspermatophyte plants due to the scarcity of DNA material. In the future it will be critical to sample genomic data from these symbionts of nonspermatophyte plants (particularly from Densosporaceae) to obtain the complete picture of Endogonales diversification and to better understand their roles in early land-plant evolution.

Evolutionary signatures of the EcM lifestyle

One of the most prominent features of the assembled Endogonaceae genomes is their large sizes, whereas the average fungal genome size ranges from 30 to 50 Mb (Mohanta & Bae, 2015). Among the sampled Mucoromycotina genomes, non-Endogonaceae genomes have an average size of 36 Mb. The closest relatives of Endogonaceae in our sampling, *Calcarisporiella* and *Bifiguratus*, have small genomes of 28 and 20 Mb, suggesting that the increase in genome size is unique to Endogonaceae. By contrast, the gene content of the Endogonaceae genomes is comparable to that of other fungi. The expansion of Endogonaceae genome size is due to their high repeat content, which is one of the traits shared by many mycorrhizal fungi and obligate plant pathogenic fungi (Martin *et al.*, 2008; Tisserant *et al.*, 2013; Dong *et al.*, 2015; Peter *et al.*, 2016).

A reduction in the number of PCWDEs is another genome signature often observed in Dikarya EcM fungi (Kohler *et al.*, 2015). As biotrophs, EcM fungi tend to release fewer digestive enzymes to avoid triggering the defense system of the hosts. Furthermore, they do not need a large repertoire of digestive enzymes

because they derive most of their C from their hosts. Our study shows that the Endogonaceae genomes have a reduced number of PCWDEs compared with their hypothetical ancestors and the majority of the other Mucoromycotina fungi (Fig. 5; Table S4). However, the degree of reduction of PCWDEs in Endogonaceae is smaller than reported for Dikarya EcM fungi. Mucoromycotina fungi, including their hypothetical ancestors, possess a small number of PCWDEs, especially those targeting crystalline cellulose (e.g. GH6, GH7, AA9 and CBM1). The high content of PCWDEs did not yet exist in the ancestors of Endogonales, and therefore the reduction in the number of PCWDEs is more nuanced compared with Dikarya EcM fungi.

The diversification of SSPs in EcM fungi has been described as part of the 'symbiosis molecular toolbox'. These molecules have received increased attention for their roles in fungus–plant interactions (Plett & Martin, 2015; Martin *et al.*, 2016). The SSPs that function as 'effector' proteins are believed to play an important role in the establishment of the mycorrhizal relationship. Genomes of Dikarya EcM fungi are enriched in SSP genes, especially compared with those of saprotrophic fungi (Pellegrin *et al.*, 2015; Kim *et al.*, 2016). In our analysis, Endogonaceae genomes do not show elevated levels of SSPs when compared with other Mucoromycotina species (Fig. 5). The SSP/secretome ratios of Endogonaceae genomes are lower than those of Basidiomycota EcM genomes sampled here and reported in other studies (Pellegrin *et al.*, 2015). By contrast, the Endogonaceae genomes have high proportions of species-specific SSPs (SSSPs) compared with that of Basidiomycota EcM fungal genomes. The SSSP/SSP ratios of the Basidiomycota EcM species typically vary between 25% and 40% (Pellegrin *et al.*, 2015; Kim *et al.*, 2016), whereas 48–68% of the Endogonaceae SSPs are species specific (Fig. 5), close to the values reported for pathogenic fungi (which are typically 40–80% SSSPs) (Pellegrin *et al.*, 2015). The combination of moderate levels of SSPs and high levels of SSSPs suggests that Endogonaceae may have taken different strategies to interact with their host plants.

Mycorrhizal evolution in Endogonaceae with reference to the trophic mode of *Endogone* sp. FLAS F-59071

The characterization of the *Endogone* sp. FLAS F-59071 genome provides additional insight into its ecology. Most members of *Endogone*, including *Endogone* sp. FLAS F-59071, have been described as saprotrophic owing to the production of sporocarps on decayed woody substrates (Berch & Fortin, 1983; Warcup, 1990). However, members of the *E. pisiformis* clade were recently documented forming ectomycorrhizas on *Quercus* (Yamamoto *et al.*, 2017). In addition, *Endogone tuberculosa* and *Endogone aggregata* are also putatively EcM (Warcup, 1990). In this study, *Endogone* sp. FLAS F-59071 is found to have a large genome with high repeat content, a genomic feature often observed in EcM fungal genomes.

Moreover, the PCWDE and overall CAZyme composition of *Endogone* sp. FLAS F-59071 do not differ from those of the EcM *Jimgerdemannia* (Tables S3–S5; Fig. S6). Reconstruction of PCWDE evolution mapped a total of 20 losses of PCWDEs onto the branch leading to the MRCA of Endogonaceae. By contrast, only three losses were mapped onto the branch leading to the MRCA of *Jimgerdemannia*. This suggests that the MRCA of Endogonaceae was likely symbiotic with plants and that the saprobic capacity of *Endogone* sp. FLAS F-59071 is not substantially different from the EcM species. This is also consistent with the observation that within both *Endogone* and *Jimgerdemannia* there are taxa that form mycorrhizal-like associations with liverworts and hornworts (Desirò *et al.*, 2017). Although less information is available for Densosporaceae, there are a large number of taxa in this group that are documented from the thalli of nonspermatophyte plants (Bidartondo *et al.*, 2011; Yamamoto *et al.*, 2015; Desirò *et al.*, 2017). The association of multiple, distantly related Endogonales with a variety of nonspermatophyte plants suggests that the ancestral ecology of Endogonales is plant symbiotic. The timing of the evolution of Endogonales (Figs 3, 4) is also congruent with the hypothesis of Field *et al.* (2015a) that Endogonales played an important role in colonization of land by plants.

Acknowledgements

Fungal specimens were generously provided by Daniel Luoma (*J. lactiflua* OSC 162217) and Gian Maria Niccolò Benucci (*J. flammicorona* GMNB39). We thank Stephen Mondo for his great patience and generous help for the submission of genome data to GenBank. We thank Quinn Hanses for providing scanning electron micrograph of *J. flammicorona*. We would like to thank the anonymous reviewers for their valuable comments and suggestions. This material is based upon work supported by the National Science Foundation (DEB-1441604 to JWS, DEB-1441715 to JES, DEB-1441677 to MES, DEB 1737898 to GB), United States Department of Agriculture NIFA MICL02416 to GB, and the French National Research Agency through the Laboratory of Excellence ARBRE (grant no. ANR-11-LABX-0002-01 to FMM). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The work conducted by the US Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the US Department of Energy under contract no. DE-AC02-05CH11231.

Author contributions

YC, AD, JES, JWS and GB planned and designed the research. YC and AD performed DNA and RNA extraction. HN, LS, KB and IVG performed genome and transcriptome sequencing and coordination. YC, AD, AL, AC and JWS conducted data analysis. YC, AD and JWS wrote the manuscript with contributions from IVG, JES, FMM, MES and GB.

ORCID

Gregory Bonito https://orcid.org/0000-0002-7262-8978
Ying Chang https://orcid.org/0000-0001-8972-1557

Alessandro Desirò https://orcid.org/0000-0003-1347-2843
Igor V. Grigoriev https://orcid.org/0000-0002-3136-8903
Francis M. Martin https://orcid.org/0000-0002-4737-3715
Matthew E. Smith https://orcid.org/0000-0002-0878-0932
Joseph W. Spatafora https://orcid.org/0000-0002-7183-1384
Jason E. Stajich https://orcid.org/0000-0002-7591-0020

References

Antony-Babu S, Deveau A, Van Nostrand JD, Zhou J, Le Tacon F, Robin C, Frey-Klett P, Uroz S. 2014. Black truffle-associated bacterial communities during the development and maturation of *Tuber melanosporum* ascocarps and putative functional roles. *Environmental Microbiology* **16**: 2831–2847.

Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M *et al.* 2004. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Research* **32**: D115–D119.

Araldi-Brondolo SJ, Spraker J, Shaffer JP, Woytenko EH, Baltrus DA, Gallery RE, Arnold AE. 2017. Bacterial endosymbionts: master modulators of fungal phenotypes. *Microbiology Spectrum* **5**: FUNK-0056-2016.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD *et al.* 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**: 455–477.

Bao W, Kojima KK, Kohany O. 2015. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**: 11.

Beaudet D, Chen ECH, Mathieu S, Yildirir G, Ndikumana S, Dalpé Y, Séguin S, Farinelli L, Stajich JE, Corradi N. 2018. Ultra-low input transcriptomics reveal the spore functional content and phylogenetic affiliations of poorly studied arbuscular mycorrhizal fungi. *DNA Research* **25**: 217–227.

Benucci GMN, Bonito GM. 2016. The truffle microbiome: species and geography effects on bacteria associated with fruiting bodies of hypogeous Pezizales. *Microbial Ecology* **72**: 4–8.

Berch SM, Fortin JA. 1983. *Endogone pisiformis*: axenic culture and associations with *Sphagnum*, *Pinus sylvestris*, *Allium cepa*, and *Allium porrum*. *Canadian Journal of Botany* **61**: 899–905.

Bidartondo MI, Read DJ, Trappe JM, Merckx V, Ligrone R, Duckett JG. 2011. The dawn of symbiosis between plants and fungi. *Biology Letters* **7**: 574.

Bonfante P, Desirò A. 2017. Who lives in a fungus? The diversity, origins and functions of fungal endobacteria living in Mucoromycota. *ISME Journal* **11**: 1727–1735.

Bonfante P, Genre A. 2010. Mechanisms underlying beneficial plant–fungus interactions in mycorrhizal symbiosis. *Nature Communications* **1**: e48.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. Beast 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10**: e1003537.

Brundrett MC, Tedersoo L. 2018. Evolutionary history of mycorrhizal symbioses and global host plant diversity. *New Phytologist* **220**: 1108–1115.

Buchfink B, Xie C, Huson DH. 2014. Fast and sensitive protein alignment using Diamond. *Nature Methods* **12**: 59–60.

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M. 2008. Maker: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* **18**: 188–196.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17**: 540–552.

Chang Y, Wang S, Sekimoto S, Aerts AL, Choi C, Clum A, LaButti KM, Lindquist EA, Ngan CY, Ohm RA *et al.* 2015. Phylogenomic analyses indicate that early fungi evolved digesting cell walls of algal ancestors of land plants. *Genome Biology and Evolution* **7**: 1590–1601.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**: e772.

Desirò A, Duckett JG, Pressel S, Villarreal JC, Bidartondo MI. 2013. Fungal symbioses in hornworts: a chequered history. *Proceedings of the Royal Society B: Biological Sciences* **280**: e20130207.

Desirò A, Faccio A, Kaech A, Bidartondo MI, Bonfante P. 2015. *Endogone*, one of the oldest plant-associated fungi, host unique Mollicutes-related endobacteria. *New Phytologist* **205**: 1464–1472.

Desirò A, Hao Z, Liber JA, Benucci GMN, Lowry D, Roberson R, Bonito G. 2018. *Mycoplasma*-related endobacteria within Mortierellomycotina fungi: diversity, distribution and functional insights into their lifestyle. *ISME Journal* **12**: 1743–1757.

Desirò A, Rimington WR, Jacob A, Vande Pol N, Smith ME, Trappe JM, Bidartondo MI, Bonito G. 2017. Multigene phylogeny of Endogonales, an early diverging lineage of fungi associated with plants. *IMA Fungus* **8**: 245–257.

Dong S, Raffaele S, Kamoun S. 2015. The two-speed genomes of filamentous pathogens: waltz with plants. *Current Opinion in Genetics & Development* **35**: 57–65.

Durand D, Halldórsson B, Vernot B. 2005. A hybrid micro–macroevolutionary approach to gene tree reconstruction. In: Miyano S, Mesirov J, Kasif S, Istrail S, Pevzner P, Waterman M, eds. *Research in computational molecular biology*. Berlin, Germany: Springer, 250–264.

Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Computational Biology* **7**: e1002195.

Edgar RC. 2010. Search and clustering orders of magnitude faster than Blast. *Bioinformatics* **26**: 2460–2461.

Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for *de novo* detection of LTR retrotransposons. *BMC Bioinformatics* **9**: e18.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* **300**: 1005–1016.

Field KJ, Pressel S, Duckett JG, Rimington WR, Bidartondo MI. 2015a. Symbiotic options for the conquest of land. *Trends in Ecology & Evolution* **30**: 477–486.

Field KJ, Rimington WR, Bidartondo MI, Allinson KE, Beerling DJ, Cameron DD, Duckett JG, Leake JR, Pressel S. 2015b. First evidence of mutualism between ancient plant lineages (Haplomitriopsida liverworts) and Mucoromycotina fungi and its response to simulated Palaeozoic changes in atmospheric CO_2. *New Phytologist* **205**: 743–756.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al.* 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology* **29**: 644–652.

van der Heijden MGA, Martin FM, Selosse M-A, Sanders IR. 2015. Mycorrhizal ecology and evolution: the past, the present, and the future. *New Phytologist* **205**: 1406–1423.

Hirose D, Degawa Y, Yamamoto K, Yamada A. 2014. *Sphaerocreas pubescens* is a member of the Mucoromycotina closely related to fungi associated with liverworts and hornworts. *Mycoscience* **55**: 221–226.

Holt C, Yandell M. 2011. Maker2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**: e491.

Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. 2007. Wolf Psort: protein localization predictor. *Nucleic Acids Research* **35**: W585–W587.

Hoysted GA, Kowal J, Jacob A, Rimington WR, Duckett JG, Pressel S, Orchard S, Ryan MH, Field KJ, Bidartondo MI. 2018. A mycorrhizal revolution. *Current Opinion in Plant Biology* **44**: 1–6.

Katoh K, Standley DM. 2013. Mafft Multiple Sequence Alignment Software Version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.

Kenrick P, Strullu-Derrien C. 2014. The origin and early evolution of roots. *Plant Physiology* **166**: 570–580.

Kim K-T, Jeon J, Choi J, Cheong K, Song H, Choi G, Kang S, Lee Y-H. 2016. Kingdom-wide analysis of fungal small secreted proteins (SSPs) reveals their potential role in host association. *Frontiers in Plant Science* **7**: e186.

Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, Canback B, Choi C, Cichocki N, Clum A. 2015. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nature Genetics* **47**: 410–415.

Krings M, Taylor TN, Dotzler N, Persichini G. 2012. Fossil fungi with suggested affinities to the Endogonaceae from the Middle Triassic of Antarctica. *Mycologia* **104**: 835–844.

Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* **305**: 567–580.

Laczny CC, Sternal T, Plugaru V, Gawron P, Atashpendar A, Margossian HH, Coronado S, van der Maaten L, Vlassis NWilmes P. 2015. VizBin – an application for reference-independent visualization and human-augmented binning of metagenomic data. *Microbiome* **3**: e1.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359.

López-Escardó D, Grau-Bové X, Guillaumet-Adkins A, Gut M, Sieracki ME, Ruiz-Trillo I. 2017. Evaluation of single-cell genomics to address evolutionary questions using three SAGs of the choanoflagellate *Monosiga brevicollis*. *Scientific Reports* **7**: e11025.

Lu DS. 2016. *Genome organization in the ectomycorrhizal truffle* Rhizopogon vesiculosus. MSc thesis, Department of Botany and Plant Pathology Oregon State University, Corvallis, OR, USA.

Martin F, Aerts A, Ahren D, Brun A, Danchin EG, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V. 2008. The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **452**: 88–92.

Martin F, Kohler A, Murat C, Veneault-Fourrey C, Hibbett DS. 2016. Unearthing the roots of ectomycorrhizal symbioses. *Nature Reviews Microbiology* **14**: 760.

Martin FM, Uroz S, Barker DG. 2017. Ancestral alliances: plant mutualistic symbioses with fungi and bacteria. *Science* **356**: eaad4501.

Mohanta TK, Bae H. 2015. The diversity of fungal genome. *Biological Procedures Online* **17**: e8.

Mondo SJ, Toomer KH, Morton JB, Lekberg Y, Pawlowska TE. 2012. Evolutionary stability in a 400-million-year-old heritable facultative mutualism. *Evolution* **66**: 2564–2576.

Mosier AC, Miller CS, Frischkorn KR, Ohm RA, Li Z, LaButti K, Lapidus A, Lipzen A, Chen C, Johnson J *et al.* 2016. Fungi contribute critical but spatially varying roles in nitrogen and carbon cycling in acid mine drainage. *Frontiers in Microbiology* **7**: e238.

Mujic AB. 2015. Symbiosis in the Pacific Ring of Fire: evolutionary-biology of *Rhizopogon* subgenus *Villosuli* as mutualists of *Pseudotsuga*. PhD thesis, Department of Botany and Plant Pathology Oregon State University, Corvallis, OR, USA.

Naito M, Desirò A, González JB, Tao G, Morton JB, Bonfante P, Pawlowska TE. 2017. '*Candidatus Moeniiplasma glomeromycotorum*', an endobacterium of arbuscular mycorrhizal fungi. *International Journal of Systematic and Evolutionary Microbiology* **67**: 1177–1184.

Naito M, Morton JB, Pawlowska TE. 2015. Minimal genomes of mycoplasma-related endobacteria are plastic and contain host-derived genes for sustained life within Glomeromycota. *Proceedings of the National Academy of Sciences, USA* **112**: 7791–7796.

Orchard S, Hilton S, Bending GD, Dickie IA, Standish RJ, Gleeson DB, Jeffery RP, Powell JR, Walker C, Bass D *et al.* 2017. Fine endophytes (*Glomus tenue*) are related to Mucoromycotina, not Glomeromycota. *New Phytologist* **213**: 481–486.

Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. 2010. How many bootstrap replicates are necessary? *Journal of Computational Biology* **17**: 337–354.

Pellegrin C, Morin E, Martin FM, Veneault-Fourrey C. 2015. Comparative analysis of secretomes from ectomycorrhizal fungi with an emphasis on small-secreted proteins. *Frontiers in Microbiology* **6**: e1278.

Peter M, Kohler A, Ohm RA, Kuo A, Krützmann J, Morin E, Arend M, Barry KW, Binder M, Choi C. 2016. Ectomycorrhizal ecology is imprinted in the genome of the dominant symbiotic fungus *Cenococcum geophilum*. *Nature Communications* **7**: e12662.

Pirozynski KA, Malloch DW. 1975. The origin of land plants: a matter of mycotrophism. *BioSystems* **6**: 153–164.

Plett JM, Martin F. 2015. Reconsidering mutualistic plant–fungal interactions through the lens of effector biology. *Current Opinion in Plant Biology* **26**: 45–50.

Price AL, Jones NC, Pevzner PA. 2005. *De novo* identification of repeat families in large genomes. *Bioinformatics* **21**: i351–i358.

Quandt CA, Beaudet D, Corsaro D, Walochnik J, Michel R, Corradi N, James TY. 2017. The genome of an intranuclear parasite, *Paramicrosporidium saccamoebae*, reveals alternative adaptations to obligate intracellular parasitism. *eLife* **6**: e29594.

Quandt CA, Kohler A, Hesse CN, Sharpton TJ, Martin F, Spatafora JW. 2015. Metagenome sequence of *Elaphomyces granulatus* from sporocarp tissue reveals Ascomycota ectomycorrhizal fingerprints of genome expansion and a Proteobacteria-rich microbiome. *Environmental Microbiology* **17**: 2952–2968.

Rimington WR, Pressel S, Duckett JG, Bidartondo MI. 2015. Fungal associations of basal vascular plants: reopening a closed book? *New Phytologist* **205**: 1394–1398.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**: 539–542.

Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**: 301–302.

Sedlar K, Kupkova K, Provaznik I. 2017. Bioinformatics strategies for taxonomy independent binning and visualization of sequences in shotgun metagenomics. *Computational and Structural Biotechnology Journal* **15**: 48–55.

Selosse M-A, Le Tacon F. 1998. The land flora: a phototroph–fungus partnership? *Trends in Ecology & Evolution* **13**: 15–20.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. Busco: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212.

Smith ME, Gryganskyi A, Bonito G, Nouhra E, Moreno-Arroyo B, Benny G. 2013. Phylogenetic analysis of the genus *Modicella* reveals an independent evolutionary origin of sporocarp-forming fungi in the Mortierellales. *Fungal Genetics and Biology* **61**: 61–68.

Spatafora JW, Chang Y, Benny GL, Lazarus K, Smith ME, Berbee ML, Bonito G, Corradi N, Grigoriev I, Gryganskyi A *et al.* 2016. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* **108**: 1028–1046.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Stolzer M, Lai H, Xu M, Sathaye D, Vernot B, Durand D. 2012. Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics* **28**: i409–i415.

Strullu-Derrien C, Kenrick P, Pressel S, Duckett JG, Rioult J-P, Strullu D-G. 2014. Fungal associations in *Horneophyton ligneri* from the Rhynie Chert (*c.* 407 million year old) closely resemble those in extant lower land plants: novel insights into ancestral plant–fungus symbioses. *New Phytologist* **203**: 964–979.

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

Taylor JW, Berbee ML. 2006. Dating divergences in the fungal tree of life: review and new analyses. *Mycologia* **98**: 838–849.

Tedersoo L, Smith ME. 2017. Ectomycorrhizal fungal lineages: detection of four new groups and notes on consistent recognition of ectomycorrhizal taxa in high-throughput sequencing studies. In: Tedersoo L, ed. *Biogeography of mycorrhizal symbiosis*. Springer International, Cham, Switzerland, 125–142.

Tisserant E, Malbreil M, Kuo A, Kohler A, Symeonidi A, Balestrini R, Charron P, Duensing N, Frei dit Frey N, Gianinazzi-Pearson V *et al.* 2013. Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proceedings of the National Academy of Sciences, USA* **110**: 20117–20122.

Torres-Cortés G, Ghignone S, Bonfante P, Schüßler A. 2015. Mosaic genome of endobacteria in arbuscular mycorrhizal fungi: transkingdom gene transfer in an

ancient mycoplasma–fungus association. *Proceedings of the National Academy of Sciences, USA* **112**: 7785–7790.

Torres-Cruz TJ, Billingsley Tobias TL, Almatruk M, Hesse CN, Kuske CR, Desirò A, Benucci GMN, Bonito G, Stajich JE, Dunlap C *et al.* 2017. *Bifiguratus adelaidae*, gen. et sp. nov., a new member of Mucoromycotina in endophytic and soil-dwelling habitats. *Mycologia* **109**: 363–378.

Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower C, Segata N. 2015. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature Methods* **12**: 902.

Walker C. 1985. Taxonomy of the Endogonaceae. In: Trappe JM, Molina R, Berch SM, eds. *Proceedings of the 6th North American Conference on Mycorrhizas*. 193–199.

Walker C, Gollotte A, Redecker D. 2018. A new genus, *Planticonsortium* (Mucoromycotina), and new combination (*P. tenue*), for the fine root endophyte, *Glomus tenue* (basionym *Rhizophagus tenuis*). *Mycorrhiza* **28**: 213–219.

Warcup JH. 1990. Taxonomy, culture and mycorrhizal associations of some zygosporic Endogonaceae. *Mycological Research* **94**: 173–178.

Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R *et al.* 2014. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Research* **42**: D581–D591.

Yamamoto K, Degawa Y, Hirose D, Fukuda M, Yamada A. 2015. Morphology and phylogeny of four *Endogone* species and *Sphaerocreas pubescens* collected in Japan. *Mycological Progress* **14**: e86.

Yamamoto K, Endo N, Degawa Y, Fukuda M, Yamada A. 2017. First detection of *Endogone* ectomycorrhizas in natural oak forests. *Mycorrhiza* **27**: 295–301.

Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* **40**: W445–W451.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Endogonaceae sporocarps and zygosporangia.

Fig. S2 Vizbin analysis showing the gene space of the sequenced Endogonaceae genomes.

Fig. S3 Flowchart for genome assembly and binning procedure used in this study.

Fig. S4 Pie charts showing the microbial composition profiles of each Endogonaceae collection.

Fig. S5 Internode namers of the species phylogeny used in Notung analysis.

Fig. S6 Heat map of carbohydrate-active enzymes (CAZmes) for the fungal proteomes sampled in this study.

Table S1 Genome accession information of the sampled genomes.

Table S2 Fossil calibrations and gene family GH28 expansion calibration used in r8s anlaysis.

Table S3 Characterization of the repetitive elements in the four sequenced Endogonaceae genomes.

Table S4 The number of PCWDE modules present in the extant species and ancestral species inferred with Notung analysis.

Table S5 The number of duplication events of PCWDE modules mapped onto branches leading to extant species and ancestral species.

Table S6 The number of loss events of PCWDE modules mapped onto branches leading to extant species and ancestral species.

Table S7 The proportion of small secreted proteins (SSPs) in total secretome proteins for the sampled fungal proteomes.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.