Leveraging single-cell genomics to expand the fungal tree of life

Steven R. Ahrendt, C. Alisha Quandt, Doina Ciobanu, Alicia Clum, Asaf Salamov, Bill Andreopoulos, Jan-Fang Cheng, Tanja Woyke, Adrian Pelin, Bernard Henriussat, Nicole K. Reynolds, Gerald L. Benny, Matthew E. Smith, Timothy Y. James and Igor V. Grigoriev

Environmental DNA surveys reveal that most fungal diversity represents uncultured species. We sequenced the genomes of eight uncultured species across the fungal tree of life using a new single-cell genomics pipeline. We show that, despite a large variation in genome and gene space recovery from each single amplified genome (SAG), ≥90% can be recovered by combining multiple SAGs. SAGs provide robust placement for early-diverging lineages and infer a diploid ancestor of fungi. Early-diverging fungi share metabolic deficiencies and show unique gene expansions correlated with parasitism and unculturability. Single-cell genomics holds great promise in exploring fungal diversity, life cycles and metabolic potential.
insights into the biology and evolutionary histories of these uncultivated species. We illustrated robust placement of novel lineages among EDF, demonstrated higher than haploid ploidy as a common characteristic of these lineages and revealed interesting gene family evolution patterns outside the Dikarya. We also highlighted common metabolic deficiencies among uncultured lineages and tested whether these deficiencies could be overcome through culturing efforts with addition of limiting reagents. Collectively, these approaches will facilitate further study on diverse and uncultured environmental eukaryotes.

Results
Single-cell pipeline successfully captures fungal genomes with high completeness. We developed and applied our single-cell pipeline (Supplementary Fig. 1) to eight diverse target species. We recovered genomes from individual cells (‘1-cell’), as well as pools of multiple cells sequenced as one library (‘10-/30-/50-/100-cells’). These individual libraries (Supplementary Table 1) were combined in separate co-assemblies (a technique routinely used in microbial single-cell projects15,24,25) to maximize completeness.

Figure 1b summarizes genome completeness (per cent CEGMA (Core Eukaryotic Genes Mapping Approach)20), assembly size and total gene content of co-assemblies and individual assemblies. Generally, CEGMA percentages increased when more cells were incorporated into a library, ranging from 2.8% (1 cell, C. protostelioides) to 96.7% (100 cell, M. bicuspidata). For all species, co-assemblies were most complete (relative to the 1-cell or 50-cell/100-cell libraries), ranging from 73.36% (P. cylindrospora) to 99.34% (D. cristalligena). Assembly sizes of the 1-cell libraries range from 0.5 Mb (C. protostelioides) to 21.1 Mb (B. helicus). The largest single-library assembly was 30.1 Mb (100 cell, B. helicus). Co-assemble sizes ranged from 10.6 Mb (C. protostelioides) to 46.5 Mb (B. helicus). Predicted gene counts ranged from 111 (1 cell, C. protostelioides) to 6,941 (100 cell, D. cristalligena) for single-library assemblies. Co-assemble gene counts ranged from 3,328 (C. protostelioides) to 12,167 (B. helicus). Overall, there is a strong positive correlation between estimated genome completeness and assembly size (Fig. 2a). Within a single taxon, the trend suggests that more cells sorted per library results in more complete libraries. With the exception of R. allomycis, >75% of reads of a given library were incorporated into an assembly (Fig. 2b).

Single-cell assemblies provide robust fungal phylogeny. To help resolve fungal phylogeny, a maximum likelihood tree was built from single-cell assemblies provide robust fungal phylogeny. A maximum likelihood tree was built from single-cell assemblies (Fig. 1a). Within a single taxon, the trend suggests that more cells sorted per library results in more complete libraries. With the exception of R. allomycis, >75% of reads of a given library were incorporated into an assembly (Fig. 2b).

Single-cell genomes indicate that higher ploidy is common in EDF. Single-cell genomes provide a unique opportunity to separate polymorphisms among cells from polymorphisms within cells. The ploidy of most EDF species is largely unknown but a few have been shown to be higher than haploid27–29. Analyses with k-mer graphs and allele frequency spectra can successfully distinguish between haploid and non-haploid organisms29–31. For C. protostelioides and R. allomycis, our results indicate haploid and triploid patterns, respectively (Fig. 3a–d). We were able to identify putative single-nucleotide polymorphisms (SNPs) in single-cell libraries of both species, but only for R. allomycis do single-cell SNPs match those identified in isolate sequencing (Fig. 3e).

Single-cell sequencing can also identify heterozygous SNPs in diploid cells27. By using C. protostelioides and R. allomycis as haploid and non-haploid models, respectively, we identified five higher-than-haploid species (Fig. 3f). Moreover, because we are able to show that most variants identified in our five non-haploid species are present in multiple libraries (Fig. 3g), our patterns are consistent with heterozygous SNPs expected in all cells of an isolate. These results show that single-cell sequencing can elucidate the ploidy of fungi and suggest that the majority of EDF are non-haploid.

We also identified cell-to-cell polymorphisms among one-cell libraries of D. cristalligena (Supplementary Fig. 2). We determined that six SNPs for a particular gene were exactly shared between two libraries (AHPZW and AHPZP), whereas one was unique to a third library (AHSAA).

Single-cell genomes highlight common deficiencies in primary metabolism of uncultured fungi. Genome sequencing of uncultured EDF allows us to explore how metabolic content has changed during the course of fungal evolution. Compared to the common ancestor of fungi (Supplementary Fig. 3), D. cristalligena shows gains in high-level metabolic categories, whereas B. helicus has mainly losses.

Examination of the primary metabolism of genomes can reveal deficiencies responsible for unculturability, ideally resulting in successful culturing through supplementing media with missing metabolites34. One challenge with predicting missing function using single-cell genomes is the presence of false negatives from missing data, which we address by searching for commonalities in missing pathways across our target taxa. The conserved pathways found in 75% of ‘free-living’ fungi, overlaid with those missing from ≥5 target fungi, reflect consistent enzymatic losses (Fig. 4a).

The absence of spermidine synthase (enzyme classification (EC) 2.5.1.16) and S-adenosylmethionine decarboxylase (EC 4.1.1.50) suggests that almost all target taxa are unable to make spermidine and/or homospermidine (Fig. 4d). Spermidine is involved in the regulation of processes such as virulence and sporulation34,35, and deficiency in polyamines leads to auxotrophy and attenuation of virulence36,37.

Another deficiency common to almost all of the target genomes is in the assimilatory sulfate reduction pathway38 (Fig. 4b). ATP sulfurylase (EC 2.7.7.4) and adenyl sulfate kinase (EC 2.7.1.25) are missing from all except R. allomycis and a culturable M. bicuspidata relative (hereafter: NNRRL YB-4993). Phosphoadenyl-sulfate reductase (EC 1.8.4.8) and sulfite reductase (NADPH) (EC 1.8.1.2) are missing from all target genomes except M. bicuspidata NNRRL YB-4993.

Metabolism of biotin is another common deficiency among target genomes (Fig. 4c). In plants, it is synthesized from dethiobiotin by biotin synthase (EC 2.8.1.6)39. This enzyme is only found in M. bicuspidata and B. helicus. Similarly, biotin–(acetyl-CoA-carboxylase) ligase (EC 6.3.4.15) is only found in M. bicuspidata, R. allomycis and S. pseudopumilaleata.

Biosynthesis of thiamine phosphate is accomplished via thiamine-phosphate synthase (EC 2.5.1.3) and hydroxyethylthiazole
kinase (EC 2.7.1.50). These enzymes are absent from all target fungi except *B. helicus*. Both enzymes are absent from *M. bicuspidata* from *Daphnia* but are found in *M. bicuspidata* NRRL YB-4993 (Fig. 4c).

Entrance into the tricarboxylic acid cycle is accomplished through citrate synthase (EC 2.3.3.1), which is found in all free-living and target fungi. This enzyme facilitates the conversion of acetyl-CoA to citrate. ATP citrate synthase (EC 2.3.3.8) facilitates the same reaction but also generates ATP in the process. It is absent from all target fungi except for the chytrids *C. protostelioides* and *B. helicus* (Fig. 4f).

Based on our primary metabolism results, we attempted preliminary media-supplementing axenic culturing experiments for three mycoparasitic taxa: *D. cristalligena, S. pseudoplumigaleata* and *P. cylindrospora*. We tested the efficacy of five supplements to produce axenic growth of these fungi and obtained mixed results. *S. pseudoplumigaleata* either did not grow or was contaminated by the host fungus. *P. cylindrospora* grew axenically but weakly on all media treatments, including the control plates, but was not able to complete its life cycle on any of the media formulations. *D. cristalligena* responded best to our experiments; this fungus grew faster and more abundantly on the media that included all five supplements when compared to all other treatments. However, even on the fully supplemented axenic media, *D. cristalligena* was not able to complete its life cycle (Supplementary Fig. 4). Further experiments will be necessary to fully characterize axenic growth of these and related mycoparasitic EDF (see the Supplementary Information for complete experimental details).
Proteinase and CAZymes reveal ecology of uncultured fungi. To explore parasitism strategies, we focused broadly on abundance patterns of enzymatic protein domains. Comparing carbohydrate-active enzymes (CAZymes; http://www.cazy.org) and proteases (MEROPS; https://www.ebi.ac.uk/merops/), we found that facultative mycoparasites have a CAZyme-to-peptidase ratio most similar to plant pathogens, whereas that for obligate mycoparasites is most similar to animal pathogens (Fig. 5). B. helicus, the only saprotroph in this data set, has the most CAZymes among our single-cell genomes, with expansions of families that target pollen polysaccharides (Supplementary Table 4). To complement these broad observations, we focused narrowly on families often associated with mycoparasitism: subtilases, metallopeptidases and chitinases.

Subtilases in Zoopagomycota. Subtilases, one of the largest clans of serine endopeptidases, are found in fungal entomopathogens, mycoparasites and plant pathogens. Recent work has characterized novel categories of serine proteases in fungi in the Dikarya. We explored subtilase abundance in EDF mycoparasites and found that D. cristalligena subtilase sequences form a group distinct from others, whereas the Zoopagomycotina (P. cylindrospora, T. sphaerospora and S. pseudoplumigaleata) sequences split roughly equally between known proteinase K-like proteins and the group formed by D. cristalligena (Fig. 6a). Figure 6b shows the domain architecture of orthologous subtilase proteins predicted among the Zoopagomycota and illustrates expansions specific to D. cristalligena, T. sphaerospora and S. pseudoplumigaleata.

Metallopeptidases in Zoopagomycota. Class M36 metallopeptidases, also known as fungalysins, hydrolyze laminins, elastin, collagen and keratin. The amphibian pathogen Batrachochytrium dendrobatidis has an expansion of this metallopeptidase, which is presumably used in the degradation of keratin-rich amphibian skin. Orthologous metallopeptidase proteins were identified in the three species of Zoopagomycotina and D. cristalligena. A phylogenetic tree (Supplementary Fig. 5) highlights their relationship to other fungalysin proteins and illustrates a unique expansion among the Zoopagomycotina.

Chitinases in EDF. Chitin is a component of the fungal cell wall. It is broken down by chitinases of the glycoside hydrolase family 18 (GH18) and GH19 families and a recently described family (AA11). The two glycoside hydrolase families do not share similarities in protein sequence, structure or mechanism of action. Furthermore, GH18 chitinases are widely distributed, whereas GH19 chitinases are described mainly from plants and act as defences to insect or fungal invaders.

There are a few known fungal representatives of GH19 chitinases, exclusively from the Cryptomycota and the Microsporidia. We also found GH19 chitinases for the first time in the Chytridiomycota and the Zoopagomycota. A small expansion was found in Rhizocosmatium globosum, a saprotrophic chytridiomycete. Putative GH19 chitinases were also identified in D. cristalligena, Linderina pennispora and Coemansia reversa (Kicxellomycotina), and in P. cylindrospora (Zoopagomycotina). Only one protein from D. cristalligena had an additional CBM19 domain (Supplementary Fig. 6). D. cristalligena also has a unique expansion of the non-catalytic CBM18 family. Finally, the mycoparasites exclusively have AA11 genes, with clear expansions in D. cristalligena and T. sphaerospora (Supplementary Table 4).

Secondary metabolite expansion in D. cristalligena. Secondary metabolites are non-essential compounds produced by fungi for various purposes, including antagonism of other microorganisms, pathogenesis and iron chelation. Few secondary metabolites have been identified from EDF. Our results support this observation (Supplementary Table 5), with most of the species containing one or fewer non-ribosomal peptide synthetases (NRPSs) or NRPS-like proteins and two or fewer polyketide synthases or polyketide synthase-like proteins. However, D. cristalligena possesses 27 NRPS genes divided between two lineage-specific expansions in two specific clades (Supplementary Fig. 7). Several of these homologous NRPS proteins also show modular synteny, suggesting multiple
duplication events within \textit{D. crystallina}. Based on phylogenetic analysis of adenylation domains (Supplementary Fig. 7), these secondary metabolites are most closely related to epipolythiodioxopiperazine toxins that disrupt cell membranes. \textit{Trichoderma} mycoparasites are known to produce several epipolythiodioxopiperazine toxins, including gliotoxin\textsuperscript{48}. Future work is necessary to determine whether these epipolythiodioxopiperazine-like expansions are related to mycoparasitism in \textit{D. crystallina}. Most of the NRPS genes reside alone on relatively short contigs (Supplementary Table 6), precluding identification of a traditional fungal secondary metabolite gene cluster.

**Hydrophobins in \textit{C. protostelioides}**. Hydrophobins are small cysteine-rich proteins involved in the development of aerial hyphae in certain filamentous fungi\textsuperscript{49}. These proteins are currently only described in the Dikarya, with no evidence of their presence in EDF\textsuperscript{46}. \textit{Caulochytrium} is the only zoosporic genus known to produce aerial stalks and sporangia reminiscent of filamentous fungi\textsuperscript{49}. We found 14 putative hydrophobins in the single-cell \textit{C. protostelioides} proteome, all of which contained the highly conserved 8-cysteine marker region. Phylogeny and hydropathy profiles both suggest that the \textit{C. protostelioides} proteins are more closely related to the group 1 proteins found in the Dikarya (Supplementary Fig. 8). Furthermore, we found one putative group 2 hydrophobin in the \textit{Mortierella elongata} proteome. These findings represent the first examples of hydrophobins outside the Dikarya.

**Discussion**

In this study, we used single-cell genomics to create near-complete assemblies of uncultured fungi. This approach allowed us to capture an estimated 73–99% of the genome in multiple-cell co-assemblies ranging from 478 to 8,398 scaffolds. Our analysis shows that genome completeness from a single cell ranges from 6% to 88%, and that combining multiple cells can considerably increase assembly completeness from 6% to 80% (worst case) and from 88% to 96.7% (best case). Co-assemblies of genome data from different single cells further increase genome recovery while losing single-cell resolution for other analyses (for example, heterozygosity).
Fig. 4 | Core metabolism and individual pathways. a, Metabolic pathway map with enzymes found in ~70% of ‘free-living’ fungi and absent in at least one of the single-cell fungi shown as a gradient: darker red shades indicate more single-cell species missing the same enzyme in a given pathway. b–f, Certain pathways with a high degree of such common losses: assimilatory sulfate reduction (sulfate to sulfide) (b), thiamine biosynthesis (HMP and TZE to thiamine phosphate) (c), spermidine synthesis (S-adenosylmethionine to spermidine) (d), biotin metabolism (dethiobiotin to biotin (and biotinyl-5′-AMP)) (e), and citrate synthesis (citrate to/from oxaloacetate and acetyl-CoA) (f). APS, adenylate sulfate; HMP, 4-amino-5-hydroxymethyl-2-methylpyrimidine; HMP-P, HMP phosphate; HMP-PP, HMP diphosphate; PAPS, phosphoadenylyl sulfate; TZE, 5-(2-hydroxyethyl)-4-methylthiazole; TZP, TZE phosphate. The coloured boxes for b–f indicate the presence (solid) or the absence (shaded) of a given enzyme in a genome: Ra, R. allomycis (red); Bh, B. helicus (blue); Cp, C. protostelioi (green); Dc, D. cristalligena (purple); Pc, P. cylindrospora (orange); Ts, T. sphaerospora (yellow); Sp, S. pseudolumigaleata (brown); Mb, M. bicuspida (pink).
The co-assemblies allowed annotation of 3,328–12,167 proteins in each species, and common orthologous proteins were used to create a phylogenetic tree, placing these uncultured lineages among their cultured counterparts with robust bootstrap support. We placed Zoopagomycotina as a sister branch to Kickxellomyctena, although with minimal (69% bootstrap) support. A characteristic uniting these subphyla is that certain taxa in both groups produce merosporangia, that is, linear sporangia with few spores. Although Dimargaris has been difficult to confidently place in ribosomal RNA gene phylogenies owing to a rapid rate of sequence evolution, here, D. cristalligena is placed with strong (100%) support as a sister to the Kickxelliales. An rRNA-based phylogeny placed B. helicus with the order Spizellomycetales and Rhizophlyctidiales but without statistical support. In our maximum likelihood tree, B. helicus groups with the Spizellomycetales representative Spizellomyces punctatus. Caulochytrium was placed in the Spizellomycetales by Barr based on ultrastructural data. However, Barr’s concept of the order has been radically reshaped by molecular phylogenetics. Until this study, no molecular phylogenetic analyses have included Caulochytrium; thus, there was no clear null hypothesis for this chytrid. Although C. protostelioiides grouped with the Chytridiales representative R. globosum (97% support), the precise relationship of C. protostelioiides to other chytrids will require additional genome sequencing.

Single-cell sequencing is uniquely capable of addressing questions in fungal genetics regarding the organization of genetic variation. Analysis of single cells allows testing of whether detected SNPs are specific to an individual or due to variation among multiple genotypes in a variable population. The species targeted by single-cell sequencing were found to correspond to two different groups: non-haploid with high levels of heterozygosity and haploid with negligible heterozygosity (Fig. 3f). Importantly, taxa that we suspected of being diploid had >10,000 SNPs shared across libraries, whereas taxa that were haploid typically had <2,000 SNPs shared across libraries. One caveat is that fungi could be genetically diploid but show low heterozygosity, as is the case for Saccharomyces spp. Surprisingly, we found that R. allomyces is triploid, which is rare in fungi (for example, in Epichloë) and usually reflects a sexually sterile condition. This observation of greater than diploidy has precedent in the microsporidia (for example, the tetraploid Nosema), and the chytrid B. dendrobatidis has diploid to tetraploid nuclear. Four of the six basal fungi are not haploid, including three of four Zoopagomyctena. The preponderance of heterozygous species at the base of the fungal tree implies a diploid (or higher ploidy) ancestor of the fungi. Beyond estimating ploidy, the single-cell methods presented here could be applied broadly to other fungal taxa where culturing has been unsuccessful (for example, certain rust fungi and ectomycorrhizae), to facilitate genetic mapping, test for genetic segregation and establish mating systems using cohorts of meiotically produced spores.

The uncultured fungi in this study have diverse phylogenetic backgrounds and nutritional strategies. Most are parasites, presumably dependent on hosts for certain nutrients, which may explain their unculturability. By mapping the proteome to primary metabolism pathways, we observed major deficiencies in the ability to synthesize the full set of amino acids and polyamines, which probably result in auxotrophies. Overall, there is a mosaic of losses of essential metabolism genes across the uncultured taxa, consistent with patterns seen in the parasitic Cryptomycota. The deficiencies in the biotin and thiamine pathways in mycoparasites are intriguing given that Syncephalis species have been successfully grown using beef liver, which contains measurable quantities of both biotin and thiamine. However, Syncephalis do not sporulate nor complete their life cycle under these conditions. Most chytrid pollen saprophytes are culturable, yet the unculturability of B. helicus may be explained, as it shares deficiencies in spermidine and sulfur metabolism. M. bicuspidata from Daphnia is similar to strain NRRL YB-4993 from brine shrimp (although they are not conspecific) and lacks 15 enzymes that are broadly linked to aspects of urea, sulfate and thiamine metabolism. Given that the majority are also missing from various other target species, this may help to explain their unculturability. R. allomyces is unlikely to be axenically cultured because it is missing a large number of genes for critical pathways, losses that potentially cannot be corrected for as seen in microsporidia.

Mycoparasitic fungi face unique challenges because they must parasitize a fungal host using fungal enzymes without disrupting their own cells. However, the specificity of this antagonism is not fully characterized. Genomic analysis of the facultative mycoparasite Trichoderma spp. revealed several key features, including gene expansions and numerous antifungal-producing secondary metabolite gene clusters. Nevertheless, many fungal lineages contain mycoparasites and there is no generalized analysis of the traits that are selected for in this ecological transition. We did not find any orthologous genes that were a signal of mycoparasitism across all taxa. In obligate mycoparasites,
such as those studied here, we found a CAZyme-to-protease ratio more similar to that of animal pathogens rather than plant pathogens. An expansion of subtilases was observed among mycoparasitic Zoopagomycota, forming a group that was distinct from known subtilase families and among subtilases from other fungi. The GH19 chitinases are exclusively restricted to EDF lineages, which suggests that this gene family is ancestral but has been lost.

Having genomes from these EDF enabled us to uncover traits that were previously only described in Dikarya. We observed a vast expansion of secondary metabolism NRPS genes unique to *D. crystalligena* among other EDF. Similarly, we provide evidence for multiple hydrophobins in *C. protostelioides*, the only member of the Chytridiomycota known to form aerial hyphal-like spore-producing structures.

Here, we show that single-cell methods can dramatically expand our understanding of fungal biology. Given the large numbers of as-yet-unsequenced EDF genomes, many of which have never been cultured, these methods help to clarify the expectations for future assembly and functional annotation of genomes from those lineages. Furthermore, given the large numbers of uncultured Dikarya, these methods can be effectively applied across the phylogeny. As the scale of single-cell genomics increases, we will be able to generate a full picture of the fungal tree of life that precisely reveals the complete extent of fungal diversity, regardless of the culturability of the individual taxa.

Fig. 6 | Subtilases. a. Network figure highlighting the relationships between fungal subtilases. The distance reflects sequence similarity. Groups marked as Li-1 through to Li-4 refer to clusters previously reported by Li et al.\(^4\), from which the data set originated. *B. helicus*; *P. cylindrospora*; *R. allomycis*; *S. pseudoplumigaleata*; *T. sphaerospora*. b. Subtilase domain architecture in Zoopagomycota genomes, for proteins in a given homologous cluster: Mycoparasites sequenced using single-cell methods from this study (labelled with *) have generally more subtilase proteins. For *S. pseudoplumigaleata* and *T. sphaerospora*, these are predominantly single-domain (PF00082) proteins. Domain architecture colours: blue, PF00082; light grey, PF02225; dark grey, PF06280. Lifestyles are indicated with simplified icons. Species abbreviations are consistent with panel a and also include: *Cc*, *Conidiobolus coronatus*; *Cr*, *C. reversa*; *Lp*, *L. pennispora*. 
Methods

Strains and sample preparation. A dual culture of parasite R. allomyces CS55 with its host fungus Allomyces sp. was established and used previously to sequence the genome of the parasite. For this study, spore suspensions of R. allomyces were obtained under optimal conditions by washing the mycelium with a 50% glycerol solution. An estimated 10^−3 to 10^−2 spores of the parasite with up to 5% of host spores were obtained. The sample was preserved in 10% sterile glycerol solution, shipped on dry ice and stored at −80°C.

A dual culture of C. protostelioidei ATCC 52082 with its host Sordaria was used to isolate parasitic zoospores at 2.5×10^6 per ml. The zoospore suspension was preserved in 10% dimethylsulfoxide with 10% FBS, shipped on dry ice and stored at −80°C. C. recurvatus was grown to a high density of cells through enrichment methods using spruce pollen in bog water. The sample was obtained from Perch Pond Fen near Old Town, Penobscot County, Maine, USA, in June 2014. This enrichment culture was filtered through a 40-μm mesh (removing pollen and sporematter). Culture filtrate was concentrated by centrifugation to about 5×10^5 zoospores per ml. The sample was preserved in 10% glycerol, shipped on ice and stored at −80°C.

M. bisporudata was isolated from an infected population of the water flea Daphnia dentifera grown under laboratory conditions. D. dentifera samples were rinsed repeatedly with deionized water. Then, insect pins were used to puncture the carapace and a micropipette was used to collect haemolymph, which contained a mixture of M. bisporudata yeast cells and ascospores. Cells were preserved in 10% glycerol at a concentration of 10^5 spores per ml and stored at −80°C. C. cristalligena RSA 488 was grown on V8 juice agar (1 small can of original V8 juice (565 g), 1.63 ml, diluted to 11 with deionized water. 3 g CaCO3 and 20 g agar) and cultured with C. recurvatus recurvatus. Spores were shipped in 10% sterile glycerol.

S. pseudoplagioglaetae Benny, S1-1 was grown on Muscor mohieri on 10% wheat germ agar (Wg10 (ref. 75)). Parasite hyphae and spores were shipped in 10% glycerol.

T. phaeorosa RSA 1356 was grown on V8 juice agar in dual culture with C. recurvatus and harvested from petri plates. The sample was preserved in 50% glycerol at −80°C. C. cylindrospora RSA 2659 was cultivated on potato dextrose agar with the host Umbeloplasia isabella. The culture was grown on many petri dishes and the spores of both the fungus and the host were removed from the culture by washing the plates with 0.2% Triton X-100. An estimated 2.5×10^5 spores per ml of parasite with host were obtained and preserved in 10% glycerol at −80°C.

All mycoparasites described above are considered obligate mycoparasites. For benchmarking purposes, we used non-single-cell approaches to sequence material gathered from genomic DNA extracted from enrichment cultures of C. protostelioidei. We also took advantage of genomic resources from M. bisporudata NRR YB-4993, a parasite of brinne shrimp12 related to but not considered conspecific with the Daphnia parasitic M. bisporudata targeted in this study, and the published genome of R. allomyces.12 These three genomic DNA isolates were used to compare with their respective single-cell genomes.

Single-cell genomics pipeline. The pipeline schema is shown in Supplementary Fig. 1. After sample collection, as described in the preceding section, individual cells were isolated using a one-step or two-step FACS (BD Influx Cell sorter) procedure. Target population enrichment in the original sample, determined by microscopy or FACS. Cells containing more than one ‘N’ , with quality scores (before trimming) averaging <8 over the read or length under 40 bp after trimming were discarded. Remaining reads were mapped to a masked version of human hg19 with Bbmap, discarding all hits over 93% identity. Reads mapping to mouse were also similarly discarded. Reads were normalized to the mean coverage across all reads, removing reads that have very high coverage (>100). Individual libraries from each species were assembled with SPAdes80 (v2.6.0 or v3.0), using the ‘--k2 ‘ and ‘--single-cell ‘ options, and k-mers sizes of 21, 33 or 55. After assembly, scaffolds of <2 kb in length were removed as, in our experience, they are phylogenetically ambiguous. Assembled contigs were also compared using BLAST to a set of contaminant databases: bacterial, mouse, human, feline and canine. Finally, we performed tetramer principle component analysis and removed any outlier contigs. This extra cautious read usage for assembly probably leads to reduced completeness and removes symbiotic occurrences but guarantees one species genome instead.

Supplementary Table 1 presents summaries of HiSeq libraries of variable number of cells for each organism. Bold libraries denote inclusion in co-assemblies. Underlined libraries denote individual annotation. To generate co-assemblies, reads from each library were extracted, pooled and co-assembled, again with SPAdes. The final co-assemblies were annotated using the JGI Annotation Pipeline. For D. cristalligena, C. protostelioidei, R. allomyces and M. bisporudata, all individual libraries were additionally annotated. For P. cylindrospora, T. sphaerospora, S. pseudoplagioglaetae and C. recurvatus, only the most complete (highest CEGMA score) 1-cell and 100-cell (1 cell and 50 cells per library) individual libraries were additionally annotated. All final annotations were loaded into MycoCosm, the JGI Fungal Genomics Resource, for public presentation and comparative analysis.

Annotation completeness. CEGMA13 as a general measure of completeness was computed twice in BUSCO. However, because the underlying data for BUSCO relies heavily on Dikarya fungi, we have observed (Supplementary Fig. 9) that BUSCO dramatically underestimates the coverage of EDF lineages, particularly within the Chytridomycota, Blastocladiomycota and Mucormycota. As such, we continue to use CEGMA metrics for this study as it focuses primarily on EDF.

Heterozygous polymorphism discovery. Paired-end reads were aligned to assembled draft genomes using BWA14 (v0.7.12-r1044) using default parameters. Variants were identified using FreeBayes15 (v1.0.2-28-g05b252) with parameters ‘−4−pooled-continuous −min-coverage 5’ . As non-haploid organisms have multiple copies of one genome per cell, single cell-genome reads were compared to a set of contaminant sequences, at both the read level (pre-assembly) and the contig level (post-assembly). All Illumina reads were run through a filtering pipeline prior to assembly. BBDDuk (https://sourceforge.net/projects/bbmap/) was used with the settings filterk = 27 and trimk = 27 to remove Illumina adapters, known Illumina artefacts, phiX and quality trim both ends to 1Q2. Resulting reads containing more than one ‘N’ , with quality scores (before trimming) averaging <8 over the read or length under 40 bp after trimming were discarded. Remaining reads were mapped to a masked version of human hg19 with Bbmap, discarding all hits over 93% identity. Reads mapping to mouse were also similarly discarded. Reads were normalized to the mean coverage across all reads, removing reads that have very high coverage (>100). Individual libraries from each species were assembled with SPAdes80 (v2.6.0 or v3.0), using the ‘−k2 ‘ and ‘−single-cell ‘ options, and k-mers sizes of 21, 33 or 55. After assembly, scaffolds of <2 kb in length were removed as, in our experience, they are phylogenetically ambiguous. Assembled contigs were also compared using BLAST to a set of contaminant databases: bacterial, mouse, human, feline and canine. Finally, we performed tetramer principle component analysis and removed any outlier contigs. This extra cautious read usage for assembly probably leads to reduced completeness and removes symbiotic occurrences but guarantees one species genome instead.
Functional analyses. Subtilase cluster analysis was achieved using CLANS\(^8\) and a list of identified protein sequences described in Li et al.\(^9\). Metallopeptidase and hydrolipidin trees were constructed using RAxML\(^8\) (v3.1b2) with a hidden Markov likelihood tree building with RAxML\(^2\). 27,751 distinct alignment patterns, 100 bootstrap replicates and using the GAMMA protein model. For all 29,255 positions in the alignment, the target genomes had a median of 14% missing, compared to the median of 2.5% missing of others.

Phylogenetic analysis. Phylogenetic analysis of identified NRPS genes, adenylation domains were mined from proteomes using HMMER\(^8\) (v3.1b2) with a hidden Markov model based on the fungal and bacterial domains identified by Budkeley and Turgeon\(^8\). MUSCLE\(^8\) (v3.8.31) was used to align extracted domains with those used for the hidden Markov model creation and gaps were removed manually.

Phylogenetic analysis. All v-all blastp\(^v\) (v2.2.26) was run using a cut-off of 1×10\(^{-5}\) on the predicted proteomes of the representative set of fungal and outgroup species provided in Supplementary Table 2. Clusters were predicted using MCL\(^8\) (v1.008) with an inflation value of 2. A python script identified clusters containing at most one gene copy per genome, allowing for up to 8 missing taxa per cluster, which resulted in 805 total clusters selected. Each cluster was aligned using MAFFT\(^8\) (v7.047) on each cluster, with a multiple sequence alignment using MAFFT\(^9\) (v7.047) on each cluster. Cluster alignments were analyzed with RAxML\(^9\) and bootstrap replicates and using the GAMMA protein model. For all 29,255 positions in the alignment, the target genomes had a median of 14% missing.

Metabolic reconstruction. Using MCL clustering\(^8\), orthologues were collected using a minimum three taxa per orthogroup, ignoring single-cell C. protoestiolides and R. allomycis in favour of the respective co-assemblies. The R package APE\(^8\) was used for ancestral-state reconstruction. Protein functions predicted using PriAM\(^8\) were mapped to each internal node of the phylogenetic tree presented in Fig. 1. Gains and losses in single-cell lineages were considered as relative to the ancestral fungal node (Supplementary Fig. 3): D. cristalligena shows noticeable gains in all high-level metabolic categories, whereas B. helicis has mainly losses.

Data availability. The co-assembled genomes and annotations of the target species are available through MycoCosm (https://genome.igi.gov/Shungfung) and Genbank using the following MycoCosm URLs and NCBI accessions, respectively: R. allomycis CF55 single cell (https://genome.igi.gov/doi/roza1_S1; QUVT00000000), B. helicis Perch Fenn single cell (https://genome.igi.gov/doi/Bluye1; QPFV00000000), C. protoestiolides ATCC 5208 single cell (https://genome.igi.gov/doi/GeniCr; QRJA00000000), P. cylinodrosa RSA 2659 single cell (https://genome.igi.gov/doi/PPcy3_1; QPFV00000000), T. sphaerospora RSA 1356 single cell (https://genome.igi.gov/doi/Thasp1; QUU000000000), S. pseudoplumagaleata Benny 571-1 single cell (https://genome.igi.gov/doi/Traprn1; QUVV0000000000), and M. bicuspitata single cell (https://genome.igi.gov/doi/Herb1; QPID00000000) and M. bicuspitata single cell (https://genome.igi.gov/doi/Mebi_Scomb; QUVR0000000000). The whole-genome sequences for the non-single-cell isolate C. protoestiolides ATCC 5208 is available through MycoCosm (https://genome.igi.gov/doi/GeniCr1) and Genbank (QAVJ00000000). The whole-genome sequences for the non-single-cell isolate R. allomycis CF55 was not determined in this study and is available through MycoCosm (https://genome.igi.gov/doi/roza1_1) and Genbank (LXTC00000000).
92. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).

93. Claudel-Renard, C., Chevalet, C., Faraut, T. & Kahn, D. Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Res.* **31**, 6633–6639 (2003).

94. Yamada, T., Letunic, I., Okuda, S., Kanehisa, M. & Bork, P. iPPath2.0: interactive pathway explorer. *Nucleic Acids Res.* **39**, W412–W415 (2011).

Acknowledgements
The work conducted by the US Department of Energy JGI, 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. The authors thank M. Duffy (University of Michigan, MI, USA) for providing the sample of *M. bicupidata*, J. Longcore (University of Maine, ME, USA) for *B. helicus* and N. Ivanova for metabolic biochemistry insights. S.R.A., I.V.G., T.Y.J. and C.A.Q. were supported by the NSF grant DEB-1354625. M.E.S., T.Y.J., N.K.R. and G.L.B. were supported by the NSF grant DEB-1441677 (Genealogy of Life (ZyGoLife) the conundrum of Kingdom Fungi).

Author contributions
S.R.A., D.C., C.A.Q., T.Y.J. and A.P. wrote the manuscript with input from T.W., I.V.G., M.E.S. and N.K.R. D.C. performed the wet-bench protocol optimization and sequencing under supervision of J.-F.C. A.C. and B.A. performed sequencing and assembly, quality-control protocols and provided substantial technical input. S.R.A. and A.S. annotated the genomes. C.A.Q., S.R.A., A.F., A.S. and B.H. performed the comparative analyses. N.K.R. performed the media supplement axenic culture experiments. G.L.B., M.E.S. and T.Y.J. provided the biological samples. T.Y.J. and I.V.G. designed and coordinated the project.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41564-018-0261-0.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to T.Y.J. or I.V.G.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access
This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s), under exclusive licence to Springer Nature Limited 2018
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  - Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
- Clearly defined error bars
  - State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

**Data collection**

- SPAdes (Bankevich et al. 2012)
- JGI Annotation Pipeline (Grigoriev et al. 2014)
Data analysis:

- CLANS (Frickey & Lupas 2004)
- BWA (Li 2009)
- FreeBayes (Garrison & Marth 2012)
- BBTools (http://jgi.doe.gov/data-and-tools/bbtools/)
- RAxML (Stamatakis 2014)
- Muscle (Edgar 2004)
- HMMER (Eddy 2011)
- MAFFT (Katoh & Standley 2013)
- MCL (Enright, Van Dongen, & Ouzounis, 2012)
- PRIAM (Claudel-Renard et al. 2003)
- APE [R package] (Paradis et al. 2004)
- IPath (Yamada et al. 2011)
- iTol (Letunic & Bork 2016)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data. All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The co-assembled genomes and annotations of the target species are available through MycoCosm (https://genome.jgi.doe.gov/fungi) and Genbank using the following MycoCosm URLs and NCBI accessions, respectively: Rozella allomycis CSF55 single-cell (https://genome.jgi.doe.gov/Rozal_SC1; QUVT00000000), B. helicus Perch Fen single-cell (https://genome.jgi.doe.gov/Blyhe1; QPFF00000000), Caulochytrium protostelioides ATCC 52028 single-cell (https://genome.jgi.doe.gov/Caupr_SComb; QUVS00000000), Dimargaris cristalligena RSA 468 single-cell (https://genome.jgi.doe.gov/DimcrSC1; QRF60000000), Piptopcephalis cylindrospora RSA 2659 single-cell (https://genome.jgi.doe.gov/Pipcyc3_1; QPFT00000000), Thamnocephalis sphaerospora RSA 1356 single-cell (https://genome.jgi.doe.gov/Thasp1; QUUV00000000), Synccephalis pseudoplumigaleata Benny 571-1 single-cell (https://genome.jgi.doe.gov/Synsp1; QUUV00000000), and Metschnikowia bicuspidata single-cell (https://genome.jgi.doe.gov/Metbi_SComb; QUVR00000000). The whole genome sequence for the non-single-cell isolate C. protostelioides ATCC 52028 is available through MycoCosm (https://genome.jgi.doe.gov/Caupr1) and Genbank (QAJV00000000). The whole genome sequence for the non-single-cell isolate of R. allomycis CSF55 was not determined in this study, and is available through MycoCosm (https://genome.jgi.doe.gov/Rozal1_1) and Genbank (ATJD00000000). The genome sequence for the non-single-cell isolate of R. allomycis CSF55 was not determined in this study, and is available through MycoCosm (https://genome.jgi.doe.gov/Rozal1_1) and Genbank (QAJV00000000). The whole genome sequences for the non-single-cell isolate of R. allomycis CSF55 was not determined in this study, and is available through MycoCosm (https://genome.jgi.doe.gov/Rozal1_1) and Genbank (ATJD00000000). The genome sequence for the non-single-cell isolate of R. allomycis CSF55 was not determined in this study, and is available through MycoCosm (https://genome.jgi.doe.gov/Rozal1_1) and Genbank (QAJV00000000). The whole genome sequences for the non-single-cell isolate of R. allomycis CSF55 was not determined in this study, and is available through MycoCosm (https://genome.jgi.doe.gov/Rozal1_1) and Genbank (ATJD00000000).

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences  ☐ Behavioural & social sciences  ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | The fungal and eukaryotic taxa used for comparative analyses were selected to best represent the breadth of fungal diversity. In general, sample sizes represent the extent of available data. |
| Data exclusions | Exclusions were limited to institutional standard quality control filtering of sequencing data. |
| Replication | For single-cell sequencing, multiple genome libraries were constructed for each species considered. Multiple individual libraries were constructed from singular individual cells. Additionally, multiple individual libraries were also constructed from pools of individual cells. Cells from a given species were isolated from the same culture. |
| Randomization | Group allocation was not a part of this study, therefore there was no randomization. |
| Blinding | Group allocation was not a part of this study, therefore there was no blinding. |

Reporting for specific materials, systems and methods
| Materials & experimental systems | Methods |
|---------------------------------|---------|
| **n/a** | **n/a** |
| - Unique biological materials | - ChIP-seq |
| - Antibodies | - Flow cytometry |
| - Eukaryotic cell lines | - MRI-based neuroimaging |
| - Palaeontology | |
| - Animals and other organisms | |
| - Human research participants | |