environmental microbiology

Environmental Microbiology (2015) **17**(8), 2952–2968 doi:10.1111/1462-2920.12840

Metagenome sequence of *Elaphomyces granulatus* from sporocarp tissue reveals Ascomycota ectomycorrhizal fingerprints of genome expansion and a *Proteobacteria*-rich microbiome

C. Alisha Quandt,[1*†] Annegret Kohler,[2] Cedar N. Hesse,[3] Thomas J. Sharpton,[4,5] Francis Martin[2] and Joseph W. Spatafora[1]

Departments of [1]Botany and Plant Pathology, [4]Microbiology and [5]Statistics, Oregon State University, Corvallis, OR 97331, USA.

[2]Institut National de la Recherché Agronomique, Centre de Nancy, Champenoux, France.

[3]Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM, USA.

Summary

Many obligate symbiotic fungi are difficult to maintain in culture, and there is a growing need for alternative approaches to obtaining tissue and subsequent genomic assemblies from such species. In this study, the genome of *Elaphomyces granulatus* was sequenced from sporocarp tissue. The genome assembly remains on many contigs, but gene space is estimated to be mostly complete. Phylogenetic analyses revealed that the *Elaphomyces* lineage is most closely related to *Talaromyces* and Trichocomaceae s.s. The genome of *E. granulatus* is reduced in carbohydrate-active enzymes, despite a large expansion in genome size, both of which are consistent with what is seen in *Tuber melanosporum*, the other sequenced ectomycorrhizal ascomycete. A large number of transposable elements are predicted in the *E. granulatus* genome, especially Gypsy-like long terminal repeats, and there has also been an expansion in helicases. The metagenome is a complex community dominated by bacteria in *Bradyrhizobiaceae*, and there is evidence to suggest that the community may be reduced in functional capacity as estimated by KEGG pathways. Through the sequencing of sporocarp tissue, this study has provided insights into *Elaphomyces* phylogenetics, genomics, metagenomics and the evolution of the ectomycorrhizal association.

Received 18 November, 2014; revised 13 February, 2015; accepted 28 February, 2015. *For correspondence. E-mail alishaq@umich.edu; Tel. (+1) 734 763 8161; Fax (+1) 734 763 0544. †Present address: Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA.

Introduction

Elaphomyces Nees (Elaphomycetaceae, Eurotiales) is an ectomycorrhizal genus of fungi with broad host associations that include both angiosperms and gymnosperms (Trappe, 1979). As the only family to include mycorrhizal taxa within class Eurotiomycetes, Elaphomycetaceae represents one of the few independent origins of the mycorrhizal symbiosis in Ascomycota (Tedersoo *et al*., 2010). Other ectomycorrhizal Ascomycota include several genera within Pezizomycetes (e.g. *Tuber*, *Otidea*, etc.) and *Cenococcum* in Dothideomycetes (Tedersoo *et al*., 2006; 2010). The only other genome sequence published from an ectomycorrhizal ascomycete is *Tuber melanosporum* (Pezizales, Pezizomycetes), the black perigord truffle (Martin *et al*., 2010). The *T. melanosporum* genome exhibited a significant expansion in size as compared with other filamentous Ascomycota [125 megabases (Mb) versus an average of 30–40 Mb], and contained a high number of transposable elements. Conversely, it was characterized by fewer protein-coding genes than a typical Ascomycota with a reduced number of genes encoding for secondary metabolites and carbohydrate-active enzymes (CAZymes) (Cantarel *et al*., 2009), including those involved in degradation of plant material. These reductions are hypothesized to be required for the ectomycorrhizal lifestyle (Martin *et al*., 2008; 2010). This large genome size and reduction in gene number was not found in *T. melanosporum*'s closest sequenced relative, *Pyronema confluens*, which is a saprobe (Traeger *et al*., 2013), raising the question as to whether these genome-scale properties of *T. melanosporum* are common to Ascomycota ectomycorrhizal taxa.

Elaphomyces is one of two genera in the family Elaphomycetaceae, Eurotiales (Dictionary of the Fungi 10th ed.; Kirk *et al*., 2008). The type species,

E. granulatus Fr., has a broad distribution across both Europe and North America, but there has been debate about whether this represents a single species or a species complex based on morphological variation (see Hawker *et al*., 1967). The genus includes more than 90 described species that collectively have a near global distribution with species described from all continents except Africa and Antarctica (Castellano *et al*., 1989; 2011; 2012). *Elaphomyces* spp. produce subglobose, hypogeous 'truffle' fruiting bodies, which have an organized outer layer of tissue called a peridium that encloses the gleba or spore-bearing tissue (Trappe, 1979). 'Truffles' have evolved independently numerous times in multiple clades of Kingdom Fungi (Hibbett and Thorn, 2001; Hosaka *et al*., 2006), presumably under selection pressure for spore dispersal by animals (Maser *et al*., 1978; Thiers, 1984). Like most mycorrhizal symbionts, *Elaphomyces* spp. are inherently difficult to culture and maintain in the laboratory, although there are reports of successful culturing of some species (Miller and Miller, 1984).

Despite striking differences in ecology and morphology, *Elaphomyces* is a member of Eurotiales, an order that mostly comprises soil-inhabiting and medicinally important moulds, including *Penicillium* and *Aspergillus*. This evolutionary relationship was first hypothesized by Korf (1973), but it was the use of molecular data that solidified the placement of *Elaphomyces* within Eurotiales (Landvik *et al*., 1996; LoBuglio *et al*., 1996). Two studies (Miller *et al*., 2001; Geiser *et al*., 2006), using varying amounts of genetic data and taxonomic sampling, provided some support for Elaphomycetaceae as an early diverging lineage within Eurotiales, but its exact relationship to the other major clades of the order remains unclear. One uniting character for the order is the production of completely enclosed ascomata, called cleistothecia, with a few exceptions (i.e. *Trichocoma* spp.). Several genome sequences are available for *Penicillium* and *Aspergillus* spp., in addition to other eurotialean genomes including *Monascus ruber*, used in rice fermentation, and *Talaromyces* spp., many of which were traditionally classified as *Penicillium* because they often possess biverticillate penicillium-like conidiophores (Samson *et al*., 2011). Sister to Eurotiales is the order Onygenales, of which many species are animal-associated either as dermatophytes (e.g. *Trichophyton* spp.), pathogens (e.g. *Coccidioides* spp.) or coprophiles (e.g. *Histoplasma* spp.). Finally, sister to both of these orders is the Coryneliales which includes plant pathogens, such as *Caliciopsis orientalis*. Within Coryneliales, both ascus morphology and ascomatal ontogeny differ from that of species in Eurotiales and Onygenales, and their placement within the fungal tree of life was debated until the use of molecular data (Inderbitzin *et al*., 2004; Geiser *et al*., 2006).

To date, only a couple of published fungal genome sequences have been generated from tissue that was not cultured (Cantu *et al*., 2011; Cissé *et al*., 2013). However, many fungi, such as *Elaphomyces*, are difficult to grow in culture, and current genome sequencing projects are biased against sampling such species and lineages. Advances in the ability to analyse and mine complex metagenomic sequences for target sequences are beginning to enable genome sequencing from sporocarp.

Recent studies are just beginning to shed light on the microbiome of fungi and their fruiting bodies (Antony-Babu *et al*., 2013; Desirò *et al*., 2014; Splivallo *et al*., 2014). The microbial community associated with the fruiting bodies of *Elaphomyces* represents an unknown aspect of its biology, but this community could play an important ecological role in the ectomycorrhizal symbiosis (Deveau *et al*., 2007; Frey-Klett *et al*., 2007; Lehr *et al*., 2007; Kurth *et al*., 2013). Mycorrhizal helper bacteria (MHB) may promote hyphal growth and formation of mycorrhizae, producing analogues of the plant growth hormone, auxin, as has been shown of bacteria in the *Streptomycetaceae* (Riedlinger *et al*., 2006), or they could function in defining the bacterial community existing in the ectomycorhizosphere through the production of antibiotics (Frey-Klett *et al*., 2011). The most dominant bacteria commonly found within the tissues of various fungi and plants include *Pseudomonadaceae*, *Rhizobiaceae* and *Streptomycetaceae*, which are known to generally interact with hosts (e.g. nitrogen fixation, anti-pathogenic secondary metabolite production, etc.), but also include pathogenic members (Crawford *et al*., 1993; Parke and Gurian-Sherman, 2001; Offre *et al*., 2008; Hayat *et al*., 2010). Bacteria are either extrahyphal or endohyphal in association with fungi; the endohyphal bacteria are often vertically transmitted, whereas it is presumed many of the extrahyphal symbionts are derived from the soil microbial community (Mondo *et al*., 2012). Endohyphal bacteria are far less characterized, but they have been identified in multiple lineages of fungi (Bertaux *et al*., 2003; Partida-Martinez *et al*., 2007; Hoffman *et al*., 2013; Desirò *et al*., 2014), including several from arbuscular mycorrhizal fungi of Glomeromycota (Bianciotto *et al*., 2003; Mondo *et al*., 2012; Desirò *et al*., 2014). Studies examining the microbiome of truffles of *Tuber* spp. have shown that both the gleba and peridium harbour complex bacterial communities (Antony-Babu *et al*., 2013), and that they are enriched in bacteria of *Bradyrhizobiaceae* (Barbieri *et al*., 2005; Antony-Babu *et al*., 2013; Gryndler *et al*., 2013). The bacteria of *Tuber* truffles have been hypothesized, among other functions, to play a role in production of truffle aroma (Barbieri *et al*., 2000).

Several questions about the biology of *Elaphomyces* remain unanswered including: (i) What similarities, if any,

exist between the *E. granulatus* genome and that of the other sequenced ectomycorrhizal ascomycete *T. melanosporum*?; (ii) How is *Elaphomyces* related to other lineages of Eurotiales?; (iii) What genes are expanded and contracted in *Elaphomyces* as compared with saprobic relatives in Eurotiales?; and (iv) What is the bacterial community composition in an *E. granulatus* truffle and is it similar to communities found in *Tuber* truffles? A draft genome of *E. granulatus* was sequenced to obtain insight into these questions. Due to the absence of an isolated culture, *E. granulatus* DNA was extracted from field-collected sporocarp tissue and subject to shotgun metagenomic sequencing. Analysis of these metagenomic data characterized the gene space of *E. granulatus*, which in turn enabled its comparison to a broad representation of eurotialean fungi, and quantified the taxonomic and functional diversity of the *Elaphomyces* microbiome.

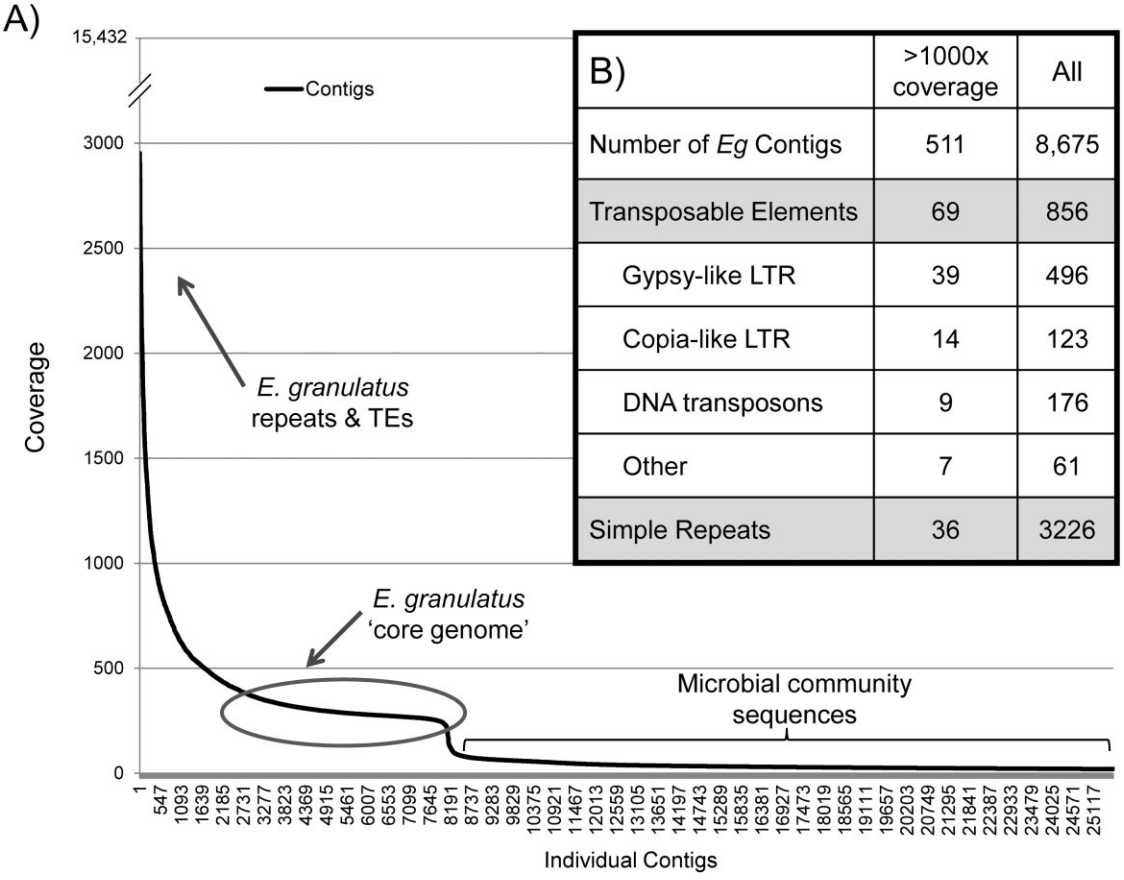
Results

Genome structure and content

The initial assembly of all ~ 321 million reads from the fruiting body resulted in 61 028 contigs with a total of 286 657 254 base pairs. Four thousand and thirty contigs had an average coverage of between 320x to 240x. These represented the core contigs of the *E. granulatus* genome (Fig. 1A), whereas the vast majority of fruiting body-associated bacterial contigs had coverage of less than 100x. Contigs with coverage > 320x were also part of the *E. granulatus* genome (with top BLAST hits to Eurotiomycetes sequences), and represent repetitive sequences (see following page).

After the conservative removal of non-*Elaphomyces* contigs, the final version of the *E. granulatus* draft genome is composed of 8675 contigs encompassing 54.2 Mb with an N50 of 11 830 bp (Table 1). There are

A)



B)

	>1000x coverage	All
Number of *Eg* Contigs	511	8,675
Transposable Elements	69	856
Gypsy-like LTR	39	496
Copia-like LTR	14	123
DNA transposons	9	176
Other	7	61
Simple Repeats	36	3226

Fig. 1. Coverages of initial contigs and their associations.
A. Contigs are plotted as individual points along the x-axis, with their respective coverages along the y axis. For visualization purposes contigs with coverage over 3000x and less than 20x have been omitted from the graph. The average coverage for most *E. granulatus* contigs is approximately 270x, while contigs with higher coverage are repetitive portions of the *E. granulatus* genome. Contigs representing the microbial community of the peridium had coverages much lower than the *E. granulatus* 'core genome.'
B. RepeatMasker annotation of repetitive sequences and transposable elements in the 460 contigs with coverage higher than 1000x.

Table 1. Genome assembly statistics and comparison with relatives in Eurotiales and the distantly related ectomycorrhizal truffle *T. melanosporum*. The lower GC contents in the *E. granulatus*/*Talaromyces* lineage are shaded.

	E. granulatus	*T. stipitatus*[a]	*T. marneffei*[b]	*P. chrysogenum*[c]	*A. fumigatus*[d]	*T. melanosporum*[e]
Size (MB)	71.5 (54.2)	35.7	28.64	32.2	29.2	124.9
Longest scaffold	90 200	5 640 589	6 407 042	6 387 817	4 896 001	2 785 000
GC %	45.6	46.1	46.7	49.0	49.5	52.0
N50 (bp)	11 801	4 363 329	3 339 384	3 889 175	3 791 214	638 000
Total scaffolds	8 675	820	452	49	55	398
Predicted Genes	8 464 (7 171)	13 252	10 638	13 671	9 916	7 496

a. *T. stipitatus* ATCC 10500, JCVI.
b. *T. marneffei* ATCC 18224, JCVI.
c. *P. chrysogenum* Wisconsin 54–1255, van den Berg *et al.*, 2008.
d. *A. fumigatus* A1163, Fedorova *et al.*, 2008.
e. *T. melanosproum* Mel28, Martin *et al.*, 2010.
Estimates of *E. granulatus* genome size and protein model number are given, with the actual size of the assembly and gene models with RNA evidence given in parentheses.

511 contigs with an average coverage of over 1000x, and 104 of these high coverage contigs were annotated as simple repetitive sequences or transposable elements (Fig. 1B), in addition to the *E. granulatus* rDNA and mitochondrial DNA. The *E. granulatus* rDNA had a median coverage of 4626x, and when compared with the 279x coverage of single copy genes (e.g. TEF and RPB1) resulted in an estimated 16 rDNA tandem repeat copy number. The contig with the single highest coverage (15 432x) is only 1350 bp, but is annotated by RepeatMasker as a Gypsy-like long terminal repeat (LTR). Using the custom repeat library generated in RepeatScout, RepeatMasker estimates that 13.16% of the assembled 54 Mb is repetitive (either simple or complex), and RepeatMasker also identified 496 Gypsy-like LTRs totalling 245 503 bp. Due to the high coverage of many repetitive sequences and transposable elements (Fig. 1), a kmer sampling method of raw reads was used to more accurately measure the genome size, which was estimated to be approximately 71 Mb. If approximately 7 Mb (13%) of the assembled 54 Mb is repetitive, and the difference (17 Mb) between assembly size (54 Mb) and estimated genome size (71 Mb) is primarily due to repetitive sequence, then repetitive sequence (24 Mb) accounts for approximately 34% of the *E. granulatus* genome.

The genome of *E. granulatus* is slightly AT rich, with an estimated GC content of 45.6% (Table 1). The reduction in GC content is driven by the non-coding regions of the genome, which is approximately 44.7% GC, whereas the coding regions have a GC content of 50%. Almost the entire mitochondrial genome lies on a single contig of 51 855 bp, and it shares a high degree of synteny with the mitochondrial assembly of *Aspergillus fumigatus* (Fig. S2). For the nuclear genome, there are 7171 protein coding genes which have protein models supported by RNA evidence, and another 1293 *ab initio* predicted proteins that had significant (e-value $\leq 1e^{-5}$) hits to one or more protein models in the Eurotiales genomes sampled in this study. Core Eukaryotic Mapping Genes Approach (CEGMA) analysis of the *E. granulatus* protein models identified 220 full-length alignments (88.71%) and 226 full or partial length alignments (91.13%) out of the 248 core eukaryotic genes, indicating that despite the metagenomic nature of the data, the assembly sufficiently captures a high proportion of the *E. granulatus* gene space.

Phylogenetic placement of Elaphomyces

Using the conservative set of *E. granulatus* protein models with RNA evidence, a total of 320 single copy orthologous clusters of proteins with 100% representation of all 31 taxa analysed were identified. This resulted in a final super alignment of 123 356 amino acid positions. The maximum likelihood analysis provided support for all of the nodes in Pezizomycotina (Maximum likelihood bootstrap proportion, MLBP = 100), except the node joining classes Pezizomycetes and Orbiliomycetes (MLBP = 59) (Fig. 2). *Elaphomyces granulatus* is reconstructed as sister to the two *Talaromyces* spp. sampled, and *M. ruber* is placed as sister to *Aspergillus* and *Penicillium* spp. The root of Eurotiales lies between the *Elaphomyces*/*Talaromyces* clade and the remaining taxa of Eurotiales.

In this analysis, Eurotiales and Onygenales are reconstructed as sister clades. The single representative of the Coryneliales, *Caliciopsis orientalis*, is sister to both of these orders, and is resolved as the earliest diverging sampled taxon within Eurotiomycetidae. The *C. orientalis* genome was sequenced as part of the Assembly the Fungal Tree of Life 2 project, and this plant pathogen has an estimated genome size of 28.2 Mb assembled onto 643 scaffolds (Table S2).

Comparative genomics

To identify potential losses in decomposition pathways indicative of the origin of the ectomycorrhizal symbiosis,

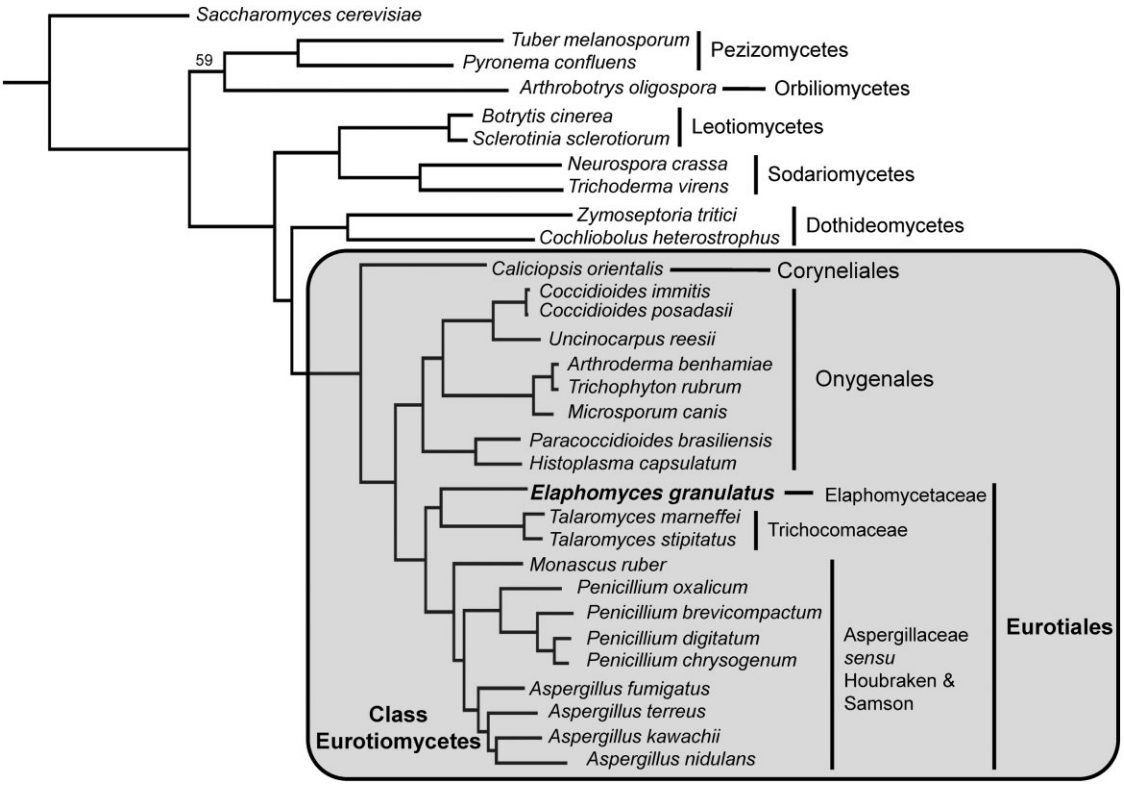


Fig. 2. Phylogeny. RAxML phylogeny created using 320 orthologous clusters from the selected Pezizomycotina taxa plus the outgroup *S. cerevisiae*. Bootstrap support for all nodes is 100, except for the node joining the Pezizomycetes and Orbiliomycetes.

CAZyme content within the *E. granulatus* genome was analysed. In contrast to sequenced Eurotiales, *E. granulatus* is reduced overall in CAZymes (Fig. 3). Of particular interest, *E. granulatus* possesses no cellulases acting on crystalline cellulose in CAZy families GH6 and GH7, which were also absent in the *T. melanosporum* genome (Martin *et al.*, 2010). This absence of GH6 and GH7 proteins is in contrast to *Talaromyces* spp. and *P. confluens*, the closest sequenced relatives of *E. granulatus* and *T. melanosporum* respectively, which both contain copies of GH6 and GH7 cellulases (Table S3). Analysis of CAZymes in Eurotiomycetes and Pezizomycetes resulted in *E. granulatus* and *T. melanosporum* clustering with the animal-associated taxa of Onygenales due to the reduction of CAZymes in these taxa (Fig. 3). *Talaromyces* spp. cluster with *Aspergillus* and most *Penicillium* spp., while *P. confluens* is grouped with taxa containing an intermediate amount of CAZymes. *Elaphomyces granulatus* is expanded in only four CAZy families compared with the other Eurotiomyces (Fig. 3), including having five copies of the endoglucanase/xyloglucanase family GH74 (the average Eurotiomycete has one and the range is one to three), and an extra copy of CBM6 which is responsible for binding either cellulose or xylose. Family GH114 is also expanded in *E. granulatus*, which has four protein models in this family of polygalactosaminidases. Another CAZyme expansion in *E. granulatus* is the presence of two GH37 proteins, trehalases, compared with the class wide average of one for Eurotiomycetes and zero in *T. melanosporum*.

The 50 097 orthologous clusters produced by the HAL pipeline at liberal Markov Cluster (MCL) analysis inflation parameter of 1.2 were analysed for *E. granulatus* lineage-specific expansions. A large group of *Elaphomyces*-expanded protein clusters were identified (Table S4), including proteins involved in regulating sulfur metabolism, ATP-dependent DNA helicases, PIF1-like helicases, mitochondrial helicases, sugar transport, nitrogen metabolism and radical S-Adenosyl methionine (SAM) domain-containing proteins; however, the majority of clusters that are expanded in copy number have unknown functions and no putative functional domains including signal peptides or transmembrane helices. By far, the most expanded clusters were helicases (Table 2). The largest protein family of helicases in *E. granulatus* contains 31 protein models (cluster 1200611), compared with a subphylum wide average of 3 and a range of zero to 24 (*Talaromyces stipitatus* has 24 helicases in this cluster). All of the proteins in this cluster have a best BLAST hit to fungal helicase-like proteins and all were

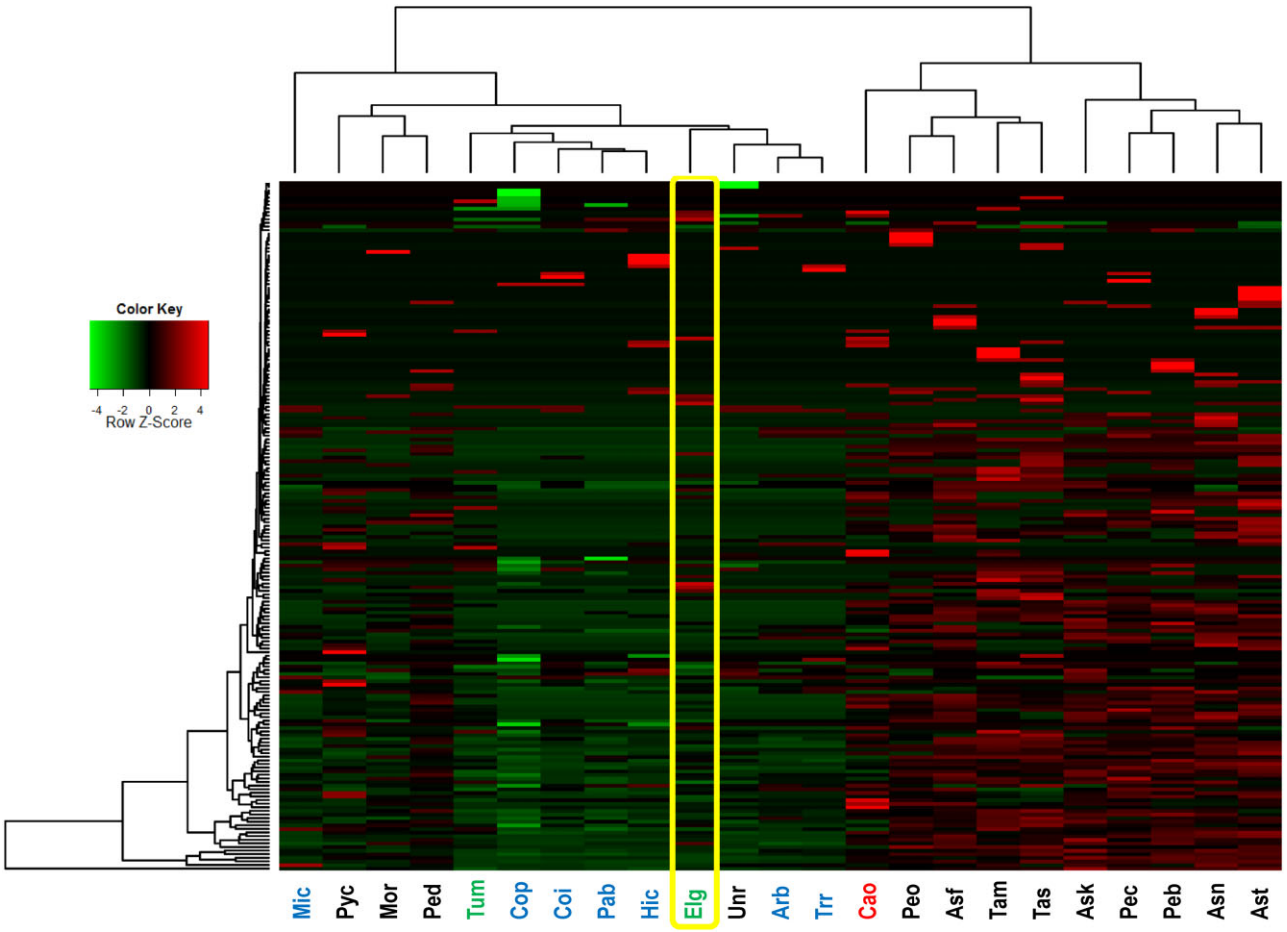


Fig. 3. Heatmap. CAZymes for all Eurotiomycetes and Pezizomycetes sampled in this study. Species names are coloured according to ecology: animal associates (blue), mycorrhizal symbionts (green), plant pathogen (red) and saprobes (black). Dendrograms calculated using Euclidean distances. Species abbreviations: Mic, *Microsporum canis*; Pyc, *Pyronema confluens*; Mor, *Monascus ruber*; Ped, *Penicillium digitatum*; Tum, *T. melanosporum;* Cop, *Coccidioides posadasii*; Coi, *Coccidioides immitis*; Pab, *Paracoccidioides brasiliensis*; Hic, *Histoplasma capsulatum*; Elg, *E. granulatus*; Unr, *Uncinocarpus reesii*; Arb, *Arthroderma benhamiae*; Trr, *Trichophyton rubrum*; Cao, *Caliciopsis orientalis*; Peo, *Penicillium oxalicum*; Asf, *Aspergillus fumigatus*; Tam, *Talaromyces marneffei*; Tas, *Talaromyces stipitatus*; Ask, *Aspergillus kawachii*; Pec, *Pencillium chrysogenum*; Peb, *Penicillium brevicompactum*; Asn, *Aspergillus nidulans*; Ast, *Aspergillus terreus*.

supported by RNA evidence; however, six of them are under 100 amino acids in length, and may represent truncated or non-functional proteins or possibly misannotated by the protein modelling program. Two of the proteins in this cluster (Egran_06015 and Egran_06016) are located on the same contig and within 2000 bp of one of the Gypsy LTRs. Another helicase cluster (120 055) contains 17 *E. granulatus* protein models, while the average number per species is seven and the range is 3 to 11 (in the opportunistic human pathogen *Histoplasma capsulatum*). Many of these helicases are also tandemly located within the genome. In addition, *E. granulatus* has five protein models in the SkpA sulfur metabolism regulator cluster (12001782), which has an average representation of one per species and a range of zero to two.

Table 2. Expanded helicase clusters in *E. granulatus* and the counts per cluster for Eurotiales and *Tuber melanosporum*.

Cluster #	Annotation	Elg	Tam	Tas	Mor	Asf	Ask	Asn	Ast	Peb	Pec	Ped	Peo	Tum	Avg.
1200611	Helicase	31	0	24	0	0	0	0	0	1	6	0	1	0	3
120055	Helicase	17	8	9	7	6	6	6	7	5	6	5	6	5	7
120067	Mito DNA helicase	7	6	9	6	6	6	7	6	4	8	6	7	3	6

Other taxa included in the analysis (see Fig. 2) have been excluded from the table due to space limitations. Species abbreviations as in Fig. 3.

Several potential secondary metabolite gene clusters were identified within the *E. granulatus* draft genome, including 3 non-ribosomal peptide synthetases (NRPSs), 18 polyketide synthase (PKS) or PKS-like genes, and 3 hybrid NRPS-PKSs. Of the NRPS core genes, one (Egran_00898) is a homologue of α-aminoadipate reductase involved in lysine biosynthesis and one (Egran_00889) is likely an intracellular siderophore synthetase that produces iron chelaters and is present in most Ascomycota. Based on the phylogeny of adenylation domains (A-domains) (Fig. S3), the third NRPS gene (Egran_01550) appears to be the product of lineage-specific duplications in *E. granulatus* as the five A-domains are all most closely related to each other. The next closest related A-domains are from an NRPS in *Penicillium chrysogenum* (Pc13g14330) and *pes1* in *A. fumigatus* (AFUB_078070) which is responsible for production of the ergoline alkaloid, fumigaclavine C (Reeves *et al*., 2006; O'Hanlon *et al*., 2012).

Elaphomyces granulatus *peridium metagenomic community analysis*

While approximately 60% (190 million reads) of the sequenced DNA was from the host, *E. granulatus*, 131 million Illumina sequences represented the metagenomic community of the peridium. This was a more than sufficient depth to characterize taxonomic diversity (Fig. S4) using the Metagenomics Rapid Annotations using Subsystems Technology (MG-RAST) server (Meyer *et al*., 2008). Twenty-four per cent of the unfiltered, untrimmed raw reads did not pass the MG-RAST quality control steps. [Read 1 (forward paired-end read) of the second library (4526251.3) had a higher percentage of reads that received a classification of unknown protein (53%) than the other reads, which had an average of 31.6% unknown protein.] Of the remaining reads, 88% were characterized as bacterial and 11% as eukaryotic, some of which could be residual host (*E. granulatus*) sequences (Fig. S5). The remaining 1% of annotated reads was classified as either virus or archaea. The MG-RAST pipeline, however, is most powerful at annotation of bacterial communities, which are the focus of the findings presented here.

At the level of class, the peridium community is dominated by *Alphaproteobacteria* accounting for over 33% of the reads sequenced (Fig. 4). *Rhizobiales* accounts for the majority of this class, but other major orders of *Alphaproteobacteria* represented in the peridium sample include *Rhodobacterales* and *Rhodospirillales*. Other classes of *Proteobacteria* were dominant portions of the peridium community as well, including *Beta-* (8.8%), *Gamma-* (7.9%) and *Deltaproteobacteria* (3.4%). Major non-*Proteobacteria* classes included *Actinobacteria* (11.8%) and *Sphingobacteria* (2.9%).

Several genera were found to be common members of the peridium community (Fig. 5). *Bradyrhizobium*, for example, was the genus most commonly attributed to the sequenced reads. The third most common genus identified, *Catenulispora*, is a genus of acidophilic members of *Actinomycetes* that have only been isolated from forest soils (Busti *et al*., 2006; Copeland *et al*., 2009). The other most common genera were *Rhodopseudomonas*, *Burkholderia* and *Streptomyces*, all of which are dominant taxa in soil communities. A large percentage of the community was identified to the candidate genera of Candidatus *Koribacter* and Candidatus *Solibacter* (both are isolates from soil and members of the phylum *Acidobacteria*) (Ward *et al*., 2009). Although none of the contigs > 100 000 bp assembled from the metagenomic data have an exact match to previously sequenced bacteria, the longest contig (> 890 Kb) assembled from the raw data had a best hit to Candidatus *Koribacter versatilis* Ellin345 (Fig. 5). On this contig, antiSMASH predicted two secondary metabolite clusters; one terpene cluster that matches a cluster in Ca. *Koribacter versatilis* and one hybrid NRPS-PKS cluster that is similar to a cluster in *Burkholderia phymatum*, a species described from legume root nodules (Moulin *et al*., 2001; 2014). This contig also contains proteins involved in the type IV secretion system. Other bacterial contigs > 100 000 bp (59 contigs) contain other secondary metabolite clusters including two PKSs, a large NRPS-lantipeptide, two bacteriocins, one phosphonate, another terpene and an ectoine that are used as osmolytes that confer resistance to salt and temperature stress. The percentage of per-taxon abundance (within the community) did not correlate with contig length, however (Fig. 5). The longest *Bradyrhizobium* contig assembled, for example, was less than 200 kb, despite accounting for more than 10% of all reads.

Functional annotation of the peridium microbiome

The use of shotgun metagenomic sequencing as opposed to targeted sequencing of ribosomal markers provides an opportunity to characterize the physiology of the community via functional annotation of the metagenome. While nearly 1.9 million reads from the *E. granulatus* metagenome received KEGG annotations, these reads were only distributed across 368 enzyme commission (ECs) numbers. This number of ECs was fewer than expected for a metagenomic sequencing. Rarefaction analysis of the *E. granulatus* metagenome indicates that this reduced functional diversity is not a product of insufficient sequence depth (Fig. S6). In particular, ECs from pathways involved in glycan biosynthesis and metabolism, secondary metabolism, lipid metabolism, and metabolism of cofactors and vitamins are reduced or not detected at all (Fig. S7). Pathways that are present in the

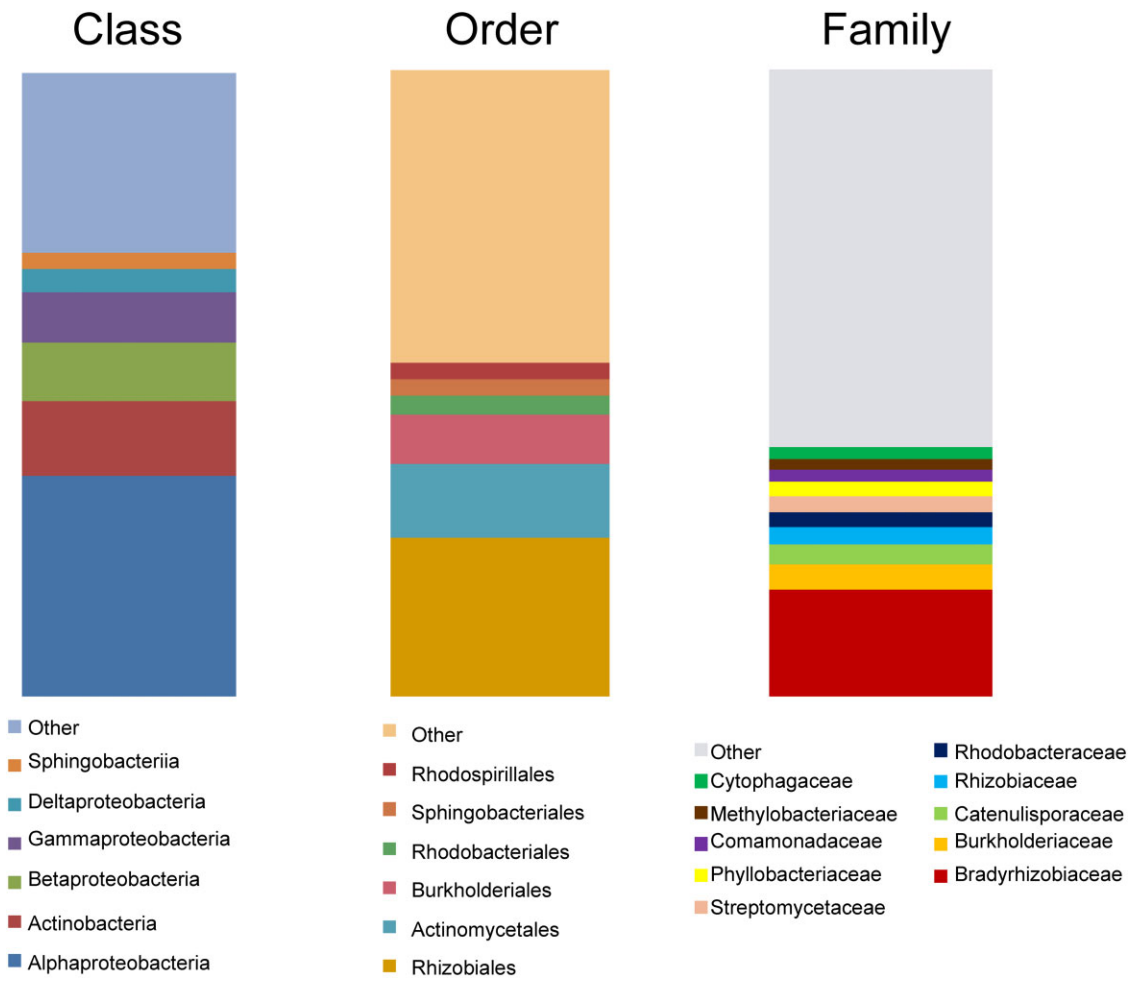


Fig. 4. Bar charts. Percentages of bacterial taxa for total annotations at given taxonomic rank found in *E. granulatus* peridium.

peridium community include nucleotide biosynthesis, the citric acid cycle, amino acid metabolism, fatty acid biosynthesis and oxidative phosphorylation.

Discussion

Fungal genome sequencing from complex sporocarp tissues

Every published fungal genome sequenced to date was generated from tissue growing in axenic culture or with the assistance of closely related reference genome sequences (Cantu *et al*., 2011; Cissé *et al*., 2013). While using tissue grown in pure culture is the most straightforward approach to generating genome sequences, it has created a significant bias against understanding genome content and organization in uncultured organisms. Many fungi cannot be readily cultured either due to obligate biotrophy or other nutritional or abiotic conditions not met by standard culture conditions. In bacteria, the analysis of metagenomes has resulted in the recovery of assembled genomes in uncultured organisms (Wrighton *et al*., 2012). This study demonstrates that analysis of a metagenome generated from *E. granulatus* sporocarp tissue reveals both the genomic content of this uncultured fungus as well as the structure and physiology of the microbiome that associates with it. This approach to the study of uncultured fungal genomes will help address systematic sampling biases in phylogenomic analyses and inform biological inquiry into the nature of obligate symbioses and the complex ecological relationships between fungi and other microorganisms. In addition, generating draft fungal genomes through metagenomic analysis may yield improvements in culturing fungi as these draft genomes can provide information that can be used to design media that are better suited to culture organisms of interest. For example, the genome sequence of one species of bacteria has also facilitated the development of culture media that is suited to pure isolation of that species (Renesto *et al*., 2003).

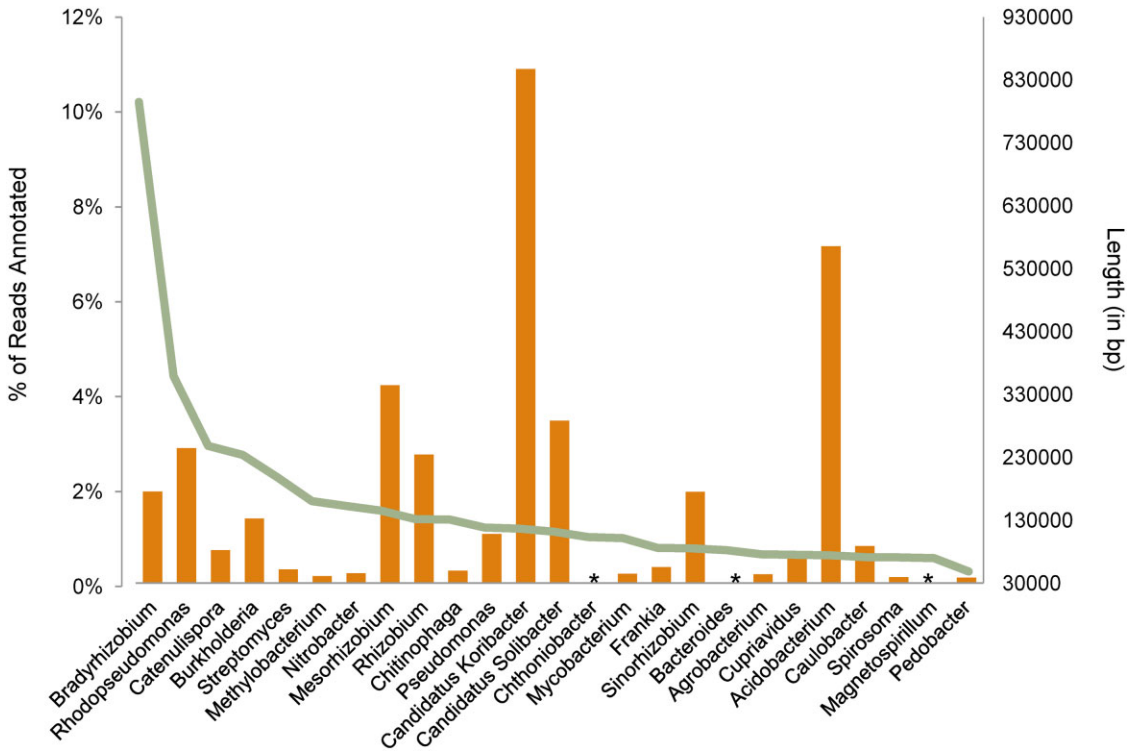


Fig. 5. Twenty-five most common genera. Abundances of the top 25 most common genera in the *E. granulatus* peridium community are represented on the left y-axis. The longest assembled contig for each of those taxa is denoted by bar charts based on the right y-axis. Asterisks denote where longest assembled contig annotated as a taxon was less than 30 000 bp.

Genome expansion and repetitive element proliferation

Elaphomyces granulatus has a highly repetitive (13–34%) and large genome (54 Mb to 71 Mb) compared with the closest sequenced relatives and any other sequenced species in Eurotiales, which range in approximate size from 24 to 37 Mb (Grigoriev *et al.*, 2014). This pattern of a relatively expanded genome size is similar to what is seen in the ectomycorrhizal truffle species, *T. melanosporum* (Martin *et al.*, 2010). *Tuber melanosporum*, like *E. granulatus*, also contains a large percentage of transposable elements, the most common of which were Gypsy-like LTRs. While there are a few Eurotiomycetes with large Gypsy content including *Paracoccidioides* spp. (*Paracoccidioides lutzii* has 442 and *Paracoccidioides* brasiliensis has 199) (Desjardins *et al.*, 2011), both *E. granulatus* and *T. melanosporum* are predicted to contain more. Since these two species are the only ectomycorrhizal ascomycete genomes sequenced to date, it is difficult to know if this is a consistent pattern for this ecology in Ascomycota. In Agaricomycotina, there is a large variation in genome size among ectomycorrhizal species [e.g. 65 Mb in *Laccaria bicolor*, 175 Mb in *Tricholoma matsutake* (MycoCosm, Grigoriev *et al.*, 2014)]. A key difference, though, is that there are fewer Ascomycota that are ectomycorrhizal and there have been fewer transitions to the mycorrhizal symbiosis than in Agaricomycotina (Hibbett and Matheny, 2009; Tedersoo *et al.*, 2010).

The low GC content (45.9%) of *E. granulatus* is a character shared with *Talaromyces* spp. (Table 1), and differs from the nucleotide content observed in *Aspergillus*, *Penicillium* and also *T. melanosporum* (Pel *et al.*, 2007; van den Berg *et al.*, 2008; Fedorova *et al.*, 2008; Martin *et al.*, 2010). When these observations are placed in a phylogenetic context, the lower GC content might be best explained as a phylogenetic character state inherited from their last common ancestor. However, the coding regions of *E. granulatus* have a higher GC content (50%) than non-coding (44.7%), suggesting that the non-coding regions are AT rich and driving the overall low genome-wide GC content. Based on this observation, genome expansions resulting from independent invasions of AT-rich repetitive elements would result in parallel divergences of genome nucleotide content. Further sampling of other taxa within the *Talaromyces* s. str. and *Elaphomyces* lineage (e.g. *Trichocoma*, *Rasamsonia*, *Thermomyces*) is necessary to differentiate between the lower GC content in the *Elaphomyces–Talaromyces* lineage versus a pattern of convergent evolution stemming from increased repetitive DNA content of the *E. granulatus* genome.

Phylogenetic relationships

This is the first study to find definitive statistical support for the sister relationship between *Talaromyces* s. str. and *Elaphomyces* (Fig. 2) and informs a number of evolutionary hypotheses for the order. Many *Talaromyces* produce a penicillium-like asexual state, and thus, it appears that the absence of asexual reproduction in Elaphomycetaceae represents a derived (or unobserved) character state, and not the ancestral state for Eurotiales. In addition, these genome scale data provide support for some of the intra-ordinal classifications advanced by Houbraken and Samson (2011). They describe and emend the large, paraphyletic family Trichocomaceae, which they restrict to *Talaromyces* and other taxa that are closely related to the type genus, *Trichocoma* (for which no genomic sequence is currently available). Traditionally, Trichocomaceae included all the eurotialean taxa included in this study's analysis except *Monascus ruber* and *E. granulatus*, members of the Monascaceae and Elaphomycetaceae respectively. Aspergillaceae sensu Houbraken and Samson (2011) includes *Aspergillus* s. str. and *Penicillium* s. str. along with *Monascus* and other related genera. Houbraken and Samson (2011) did not include any data from Elaphomycetaceae in their study, but the few studies (with very limited taxon and molecular sampling) to include data from *Trichocoma* and Elaphomycetaceae (Geiser *et al*., 2006; Tedersoo *et al*., 2010; Morgenstern *et al*., 2012) show a close relationship between these genera.

As anticipated, the sister relationship between Eurotiales and Onygenales (Geiser *et al*., 2006; Sharpton *et al*., 2009; Desjardins *et al*., 2011) is reconfirmed in the phylogenomic analysis presented here. Within Onygenales, these results are congruent with findings from other genome scale studies, specifically, the sister relationship between Onygenaceae and Ajellomyceteaceae (Desjardins *et al*., 2011). This is the first study to analyse genomic data for the plant pathogen, *C. orientalis*. Supporting what was seen in multigene phylogenies (Geiser *et al*., 2006), the concatenated genome scale analysis resolves the Coryneliales as the earliest diverging lineage of the Eurotiomycetidae, one of two subclasses of Eurotiomycetes.

Comparative genomics

The significant reduction in CAZymes in *E. granulatus* as compared with closely related taxa is similar to what was observed in *T. melanosporum*, as well as the ectomycorrhizal basidiomycetes *L. bicolor* and *Amanita* spp. (Martin *et al*., 2008; 2010; Wolfe *et al*., 2012). Many of the CAZymes that could be involved in degradation of plant cell wall materials including cellulases, hemicellulases and xylases (Cantarel *et al*., 2009) are reduced or absent in the *E. granulatus* genome (Table S3). The reduction in these genes is proposed to facilitate symbiosis with plant hosts, by not eliciting a plant immune response. This reduction may also be due to the fungus's direct access to simple carbohydrates from the host plant due to the mycorrhizal interaction. Reduction of these same plant tissue-degrading CAZymes is also seen in taxa of Onygenales (Fig. 3), as these species are animal-associated fungi. *Caliciopsis orientalis*, the only plant pathogen sampled within Eurotiomycetes, is quantitatively similar in plant-degrading CAZymes to the saprobic taxa of Eurotiales (e.g. *Aspergillus*, *Penicillium*). The expansion in trehalases of CAZy family GH37 is perplexing, as most plants do not produce trehalose, but it is commonly produced by most fungi. Trehalose is a disaccharide of two glucose units and functions in prevention of desiccation, as a signalling compound, and as a source of energy (Elbein *et al*., 2003). Fungi absorb monosaccharides like glucose, the primary form of carbon transferred from plants to mycorrhizal fungi, and convert them to trehalose, which has decreased permeability across the cell membrane and has been proposed as a carbon sink within mycorrhizal fungi (López *et al*., 2007; Wiemken, 2007). Previous studies have proposed a connection between MHB and trehalose (Duponnois and Kisa, 2006), but whether this expansion of trehalases is related to the transition to the mycorrhizal lifestyle of *Elaphomyces* or its interaction with its microbiome will require further study. Trehalases could also be involved in the need to quickly manage the osmotically active trehalose in the moisture variable soil environment in which *E. granulatus* grows.

The largest lineage specific expansions in the *E. granulatus* genome are numerous helicase-type protein families (Table 2). Helicases are responsible for separating the strands of nucleic acids (DNA and RNA) during DNA replication, transcription of RNA and many other processes (Patel and Donmez, 2006). It is possible that these are involved in the proliferation of transposable elements and the expansion of the *E. granulatus* genome, although only one of these helicases is located next to one of the many Gypsy LTRs in the genome.

A large number of secondary metabolites have been described from species of *Aspergillus*, *Penicillium* and related taxa using both genome mining and traditional chemical isolation approaches (Keller *et al*., 2005; Nierman *et al*., 2005; Khaldi *et al*., 2010; Inglis *et al*., 2013). In contrast, the genome of *E. granulatus* is depauperate in, though not without, secondary metabolic core genes compared with other eurotialean taxa. There are several predicted PKS clusters in the *E. granulatus* genome, but relatively few NRPS cluster compared with other Eurotiales (Khaldi *et al*., 2010). Because NRPSs are large and modular, the few predicted clusters may be the

result of short read sequence technology; however, this phenomenon was not observed in other genome sequencing projects (e.g. C. A. Quandt, unpublished). Limited presence of secondary metabolism was seen in the genome of *T. melanosporum* (Martin *et al*., 2010), in the Onygenales (Khaldi *et al*., 2010) and in the obligate biotrophic plant pathogen, *Blumeria graminis* (Spanu *et al*., 2010). One known function of secondary metabolites is as pathogenicity factors (Fox and Howlett, 2008), and as *E. granulatus* is a mycorrhizal species, it may not require a large repertoire of secondary metabolites as that of the other, saprobic eurotialean taxa.

The E. granulatus *peridium microbiome*

Bacteria accounted for the majority of sequences annotated in the microbiome (88%). Due to two factors, the authors have focused on the bacterial components here. One, the annotation pipeline used, MG-RAST, does not analyse the eukaryotic coding regions and is therefore not recommended for in depth analysis of the eukaryotic communities (Meyer *et al*., 2008). Finally, definitively assigning short read sequence data to fine-scale taxonomic classifications in fungi is inherently difficult without a reference genome for *E. granulatus*, making delineation between host genome and eukaryotic microbiome tenuous.

Proteobacteria dominated the sequences of the *E. granulatus* microbiome, especially those of the class *Alphaproteobacteria* which accounted for more than 33% of the sequences annotated (Fig. 4). Other abundant non-proteobacteria classes included *Actinobacteria* and *Sphingobacteria*. The largest order of sequences present was *Rhizobiales*, with the greatest proportion of these belonging to *Bradyrhizobiaceae*.

The dominance of *Bradyrhizobium* in the *E. granulatus* peridium (Fig. 5) mirrors results from studies examining the truffles of *Tuber* spp. (Barbieri *et al*., 2005; 2007) and one study using high-throughput data examining *T. melanosporum* fruiting body communities (Antony-Babu *et al*., 2013). This is particularly interesting given the phylogenetic distance between *Tuber* and *Elaphomyces*, and could be the result of a similar, suitable niche for *Bradyrhizobium* provided by the truffles. In contrast, *Catenulispora* has not been found as a common component of the communities associated with other fungal fruiting bodies. Since adjacent soil cores were not sampled in addition to the truffle, it is unknown if *Catenulispora* is also present in similar abundance in the surrounding soil, or if this species is more abundant within the *E. granulatus* peridium. *Rhodopseudomonas*, *Burkholderia* and *Streptomyces* were other dominant taxa in the community. Some species of *Streptomyces* and the *Burkholderiales* have been identified as MHB in numerous systems (Frey-Klett and Garbaye, 2005). These findings suggest that these bacteria may be selected for in the intra-fruit body niche. *Acidobacteria*, Candidatus *Koribacter versatilis* and Candidatus *Solibacter usitatus* were other abundant members or the community, and several of the longest bacterial contigs assembled had hits to these taxa (Fig. 5). These are candidate species that are difficult to grow in culture, but due to their common presence in soil have recently been targeted for genome sequencing (Ward *et al*., 2009).

The abundance of reads annotated as a particular taxon did not correlate with the ability to assemble genomes as measured by longest contig (Fig. 5). This was unexpected and three factors could be contributing to this. One, the community could contain several related taxa (e.g. many species from a genus) that are diverged to an extent that assembly into longer contigs is compromised. Two, large genome size and therefore low coverage could be inhibiting assembly of long contigs. This could explain the relatively poor assembly of the most abundant genus, *Bradyrhizobium*, species of which can have relatively large genome sizes (> 9 Mb) (Boussau *et al*., 2004). Three, some of the genomes could be repetitive, limiting the ability to assemble long contigs.

Comparison of the functional diversity based on KEGG annotations resulted in few EC numbers being annotated from the *E. granulatus* peridium sequences (Fig. S7), and rarefaction analyses of these annotations suggested this was not due to under sampling (Fig. S6). It is possible that there are novel, currently undescribed pathways that exist in this microbiome; however, based on the described, reference KEGG pathways, the *Elaphomyces* peridium community appears to be functionally reduced. This apparent reduction could be indicative of a microbiome that relies on its *Elaphomyces* host for some metabolic functions (e.g. nutrition), or that is specialized in what metabolic functions it can provide to its host fungus.

Conclusions

Given the vast number of fungi that are difficult if not impossible to grow in culture, it is important to develop culture independent techniques for genome – and therefore metagenome – sequencing and analysis. The ectomycorrhizal truffle *E. granulatus* was successfully sequenced from sporocarp tissue resulting in a robust assessment of the majority of gene content. Bioinformatic analyses support that there are many more similarities in genome size, structure and gene content between *E. granulatus* and *T. melanosporum* than between either of these and non-mycorrhizal species. When compared with their closest phylogenetic relatives, both genomes are characterized by a reduction in overall gene content, especially those associated with secondary metabolism

and CAZymes, and both genomes are characterized by an increase in overall genome size due to the presence of a high number of Gypsy-like LTRs. These findings suggest a similar pattern of evolution between these independent lineages (Eurotiales and Pezizales) of ectomycorrhizal symbionts of Ascomycota. A diverse but distinct community of bacteria, which is dominated by Proteobacteria, exists within the peridium of *E. granulatus*. Bacterial sequences of the metagenome had best hits to genera of Rhizobiales, Actinomycetales and Burkholderiales, but the number of sequences per taxon did not correlate with the longest assembled contig. The ability to assemble longer bacterial contigs was interpreted as some genera of the community being represented by fewer taxa, smaller genomes or both. Functional annotation of this community indicates it may be reduced in function, but further work is needed to understand how this community may contribute to the complex symbiosis between *E. granulatus* and its plant hosts, and if any species have evolved as endohyphal symbionts. This study provides a framework on which to further examine the evolution, symbiosis and microbiome of one of the most broadly distributed and evolutionarily unique lineages of ectomycorrhizal fungi.

Experimental procedures

Tissue extraction and nucleic acid preparation

A single collection of several *E. granulatus* fruiting bodies (OSC 145934) was made from Cummins Creek area of Siuslaw National Forest, Lane County, Oregon, (44°16′03.9″N, 124°5′59.8″W) and transferred to the laboratory for nucleic acid extraction. The largest of the fruiting bodies was cleaned of debris and surface sterilized for 30 s using 85% ethanol and a sterilized spatula was used to physically remove the warty outer peridium. The fruiting body was then cut in half and as much of the gleba as possible was removed from the peridium using a sterilized spatula. The gleba was not used in this study because it is very difficult to obtain DNA and RNA from *Elaphomyces* spores as they are highly ornamented and hydrophobic. The entire peridium was lyophilized and half of the peridium was ground in liquid nitrogen with a mortar and pestle. DNA was extracted from ground tissue using a Qiagen Plant DNeasy kit using the 'fungi' protocol without modifications. Paired-end Illumina libraries were then constructed with ~ 5 μg DNA using New England Biomedicals NEBNext kits and protocols, with gel-selected small insert sizes of 322 and 354 bp, from genomic DNA prepared from two separate extractions. Each library was sequenced for 101 cycles for each paired-end on one full lane of the Illumina HiSeq2000 at the Oregon State University Center for Genome Research and Biocomputing: the 322 bp library with Illumina version 2 chemistry and the 354 bp library with version 3 chemistry. RNA was extracted using the Qiagen Plant RNeasy kit and prepared for sequencing and sequenced at Beckman Coulter Genomics for improvement of gene model annotation.

Bioinformatic analyses

Reads were initially assembled and filtered for quality in CLC Genomics Workbench v. 6.5.1. BLASTx searches of contigs to the NCBI database initially identified several most commonly encountered bacterial taxa from the fruit body microbiome, and the subsequent methods were employed to separate bacterial data from the target genome and are summarized in Fig. S1. A database of the 35 most commonly hit genera of bacteria was created (Table S1) and raw reads were aligned to this custom database using Bowtie 2 v. 2.0.6 (Langmead and Salzberg, 2012) with default alignment settings. Reads that aligned to the bacterial database were excluded, and the remaining reads were reassembled in CLC Genomics Workbench as before. Subsequent BLASTx revealed that many bacterial sequences had not been removed due to significant divergence from sequenced reference strains. A binning procedure based on e-values ($\leq 1e^{-5}$) was then used to separate contigs into three categories: (i) those contigs with hits to sequenced Eurotiomycetes genomes available on the MycoCosm website (Grigoriev *et al*., 2014), (ii) contigs with hits to the NCBI bacterial database, and (iii) those contigs without hits to either. This third category was sorted based on coverage, either greater or less than 200x coverage. Using the contigs in category one and category three which had > 200x coverage, *ab initio* protein modelling was performed using AUGUSTUS (Stanke *et al*., 2006) with *A. fumigatus* as the reference species. Many (14 760) of the 27 436 predicted protein models were found to be bacterial using BLASTx. Manual curation of the contigs containing these bacterial hits was completed to mitigate the inclusion of chimeric contigs in the assembly. Most of these contigs with matches to bacteria contained proteins which had matches exclusively to bacterial proteins and none to Eurotiomycetes proteins, and these were deemed non-target and excluded from the genome assembly. Contigs containing one protein with bacterial matches and one or (as was the most commonly occurring case) multiple proteins with matches to Eurotiomycetes were kept in the final assembly.

Genome size estimation was performed using the kmer counting software Jellyfish v 2.0 with a kmer size of 21 (Marçais and Kingsford, 2011) and the equation presented in Li and colleagues (2010). The CEGMA was used to estimate the completeness of the *E. granulatus* genome (Parra *et al*., 2007; 2009). Whole scaffold alignments were performed in Mauve (Darling *et al*., 2010) with default progressive Mauve alignment settings. Candidate transposable elements were binned into broad classes using RepeatMasker v 3.2.8 with 'fungi' as the reference database and other default settings (Smit and Green, 1996).

Quality filtering of RNA sequences was performed based on Illumina quality flagging. Trimming of the first five and last 15 bp of RNA sequences was performed using the fastx toolkit (Gordon, 2011), and RNA sequences were assembled in Trinity (Grabherr *et al*., 2011) using default settings. The Trinity RNA assembly and the final *E. granulatus* genome assembly were imported into the Maker annotation pipeline (Cantarel *et al*., 2008) for protein model prediction. For use in the Maker pipeline, RepeatMasker v 3.2.8 was used to identify repetitive elements (Smit and Green, 1996), and a custom repeat library was created and provided to Maker using

RepeatScout v 1.0.3 with an lmer length of 14. Other filtering scripts distributed with the RepeatScout package were used to filter out low complexity and tandem sequences and those repeated no more than three times (Price *et al*., 2005). Other inputs into Maker included an hidden Markov model (HMM) created in Genemark-ES v 2.0 (Ter-Hovhannisyan *et al*., 2008) trained on the *E. granulatus* assembly, protein models from all Eurotiales genomes used in this study and *ab initio* predictions from AUGUSTUS (Stanke *et al*., 2006). Non-overlapping ab initio protein models were aligned using BLAST against a custom database of all the protein models of all Eurotiomycete genomes included in Fig. 2. Any of these protein models with a significant hit (≤ 1e-5) were included in the final protein set. All genome scaffolds and gene models were submitted to Genbank (NCBI BioProject ID 248240, *E. granulatus*).

Selection of taxa used for phylogenomic analysis was based on breadth of Pezizomycotina diversity and depth of diversity from publically available Eurotiomycetes. The phylogenomic pipeline Hal (Robbertse *et al*., 2011) was used to identify single copy orthologous clusters of proteins for use in phylogenetic analyses. Briefly, orthologous clusters of proteins were identified in MCL (Enright *et al*., 2002) across inflation parameters 1.2, 3 and 5. Orthologous clusters were filtered for retention of clusters with one sequence per genome and removal of any redundant clusters. The resulting unique, single-copy orthologous clusters of proteins were aligned in MUSCLE (Edgar, 2004) with default settings; poorly aligned regions were identified using Gblocks (Talavera and Castresana, 2007; gap removal setting = liberal) and excluded from subsequent analyses. Aligned and trimmed alignments were concatenated into a super-alignment and maximum likelihood phylogenetic analysis was performed in RAxML v 7.2.6 using the Gamma model of rate heterogeneity and the WAG substitution matrix with 100 bootstrap replicates (Stamatakis, 2006).

Hmmer3 package v 2.4 (Eddy, 2011) was used to create a custom HMM based on adenylation domains from a wide variety of characterized NRPSs (Bushley and Turgeon, 2010) and then used to identify adenylation domains of NRPS-like genes and clusters. Manual curation of these genes and clusters was completed using the NCBI database and pFAM domain annotation (Punta *et al*., 2012). Prediction of PKSs and other types of secondary metabolite genes and clusters was performed using antiSMASH (Blin *et al*., 2013) and SMURF (Khaldi *et al*., 2010). HMMs from dbCAN were used to annotate CAZymes in the *E. granulatus* genome (Yin *et al*., 2012). Heatmap clustering in R was based on default Euclidean distances as a part of the heatmap.2 program.

Bowtie 2 v. 2.0.6 alignment (Langmead and Salzberg, 2012) of raw reads to the *E. granulatus*-specific contigs was used to identify those reads that did not align and were used for downstream analysis of the microbial community. These more than 131 million reads were uploaded onto the MG-RAST webserver version 3.3.9 (Meyer *et al*., 2008). Because the read length was not long enough to assemble the paired reads across the insert size, paired-end reads were uploaded as separate files to MG-RAST server (MG-RAST IDs 4526250.3, 4526252.3, 4526251.3 and 4526253.3). MG-RAST is a web-based tool for analysing metagenomic sequence data which assigns taxonomic and functional ranks (where possible) to all sequences uploaded, regardless of whether the DNA sequenced was obtained from a particular portion of the genome (e.g. rDNA, mtDNA, etc.). All downstream analyses of microbial community were performed using data output from MG-RAST annotations of the data, and the M5NR non-redundant database was selected. Rarefaction of EC annotations was conducted in the Vegan program in R (Dixon, 2003).

Acknowledgements

The authors acknowledge the contributions made by Mark Desanko and Alija Mujic. We also thank the input and technical support from Aurelie Deveau, Emmanuelle Morin and Claude Murat at INRA. For the *Monascus ruber* genome, we acknowledge the contributions of Igor Grigoriev and Kerrie Berry at the Joint Genomes Institute of the US Department of Energy, and Kerry O'Donnell and Stacy Sink at the USDA-ARS in Peoria, IL. Funding was provided from the following sources: NSF grants DEB-0732993 to JWS, and the Laboratory of Excellence Advanced Research on the Biology of Tree and Forest Ecosystems (ANR-11-LABX-0002-01) to FM. A Laboratory of Excellence Visiting Scientist Award and an NSF Graduate Research Fellowship supported CAQ.

References

Antony-Babu, S., Deveau, A., Van Nostrand, J.D., Zhou, J., Tacon, F.L., Robin, C., *et al*. (2013) Black truffle-associated bacterial communities during the development and maturation of *Tuber melanosporum* ascocarps and putative functional roles. *Environ Microbiol* **16:** 2831–2847.

Barbieri, E., Potenza, L., Rossi, I., Sisti, D., Giomaro, G., Rossetti, S., *et al*. (2000) Phylogenetic characterization and in situ detection of a Cytophaga-Flexibacter-Bacteroides phylogroup bacterium in *Tuber borchii* Vittad ectomycorrhizal mycelium. *Appl Environ Microbiol* **66:** 5035–5042.

Barbieri, E., Bertini, L., Rossi, I., Ceccaroli, P., Saltarelli, R., Guidi, C., *et al*. (2005) New evidence for bacterial diversity in the ascoma of the ectomycorrhizal fungus *Tuber borchii* Vittad. *FEMS Microbiol Lett* **247:** 23–35.

Barbieri, E., Guidl, C., Bertaux, J., Frey-Klett, P., Garbaye, J., Ceccaroli, P., *et al*. (2007) Occurrence and diversity of bacterial communities in *Tuber magnatum* during truffle maturation. *Environ Microbiol* **9:** 2234–2246.

van den Berg, M.A., Albang, R., Albermann, K., Badger, J.H., Daran, J.-M., Driessen, A.J.M., *et al*. (2008) Genome sequencing and analysis of the filamentous fungus *Penicillium chrysogenum*. *Nat Biotechnol* **26:** 1161–1168.

Bertaux, J., Schmid, M., Chemidlin Prevost-Boure, N., Churin, J.L., Hartmann, A., Garbaye, J., and Frey-Klett, P. (2003) In situ identification of intracellular bacteria related to *Paenibacillus* spp. in the mycelium of the ectomycorrhizal fungus *Laccaria bicolor* S238N. *Appl Environ Microbiol* **69:** 4243–4248.

Bianciotto, V., Lumini, E., Bonfante, P., and Vandamme, P. (2003) Candidatus '*Glomeribacter gigasporarum*' gen. nov., sp. nov., an endosymbiont of arbuscular mycorrhizal fungi. *Int J Syst Evol Microbiol* **53:** 121–124.

Blin, K., Medema, M.H., Kazempour, D., Fischbach, M.A., Breitling, R., Takano, E., and Weber, T. (2013) antiSMASH 2.0 – a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res* **41:** W204–W212.

Boussau, B., Karlberg, E.O., Frank, A.C., Legault, B.A., and Andersson, S.G.E. (2004) Computational inference of scenarios for α-proteobacterial genome evolution. *PNAS* **101:** 9722–9727.

Bushley, K.E., and Turgeon, B.G. (2010) Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. *BMC Evol Biol* **10:** 26.

Busti, E., Cavaletti, L., Monciardini, P., Schumann, P., Rohde, M., Sosi, M., and Donadio, S. (2006) *Catenulispora acidiphila* gen. nov., sp. nov., a novel, mycelium-forming actinomycete, and proposal of Catenulisporaceae fam. nov. *Int J Syst Evol Microbiol* **56:** 1741–1746.

Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., *et al*. (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* **18:** 188–196.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* **37:** D233–D238.

Cantu, D., Govindarajulu, M., Kozik, A., Wang, M., Chen, X., Kojima, K.K., *et al*. (2011) Next generation sequencing provides rapid access to the genome of *Puccinia striiformis* f. sp. *tritici*, the causal agent of wheat stripe rust. *PLoS ONE* **6:** e24230.

Castellano, M.A., Trappe, J.M., Maser, Z., and Maser, C. (1989) *Key to Spores of the Genera of Hypogeous Fungi of North Temperate Forests with Special Reference to Animal Mycophagy*. Arcata, CA, USA: Mad River Press.

Castellano, M.A., Trappe, J.M., and Vernes, K. (2011) Australian species of *Elaphomyces* (Elaphomycetaceae, Eurotiales, Ascomycota). *Aust Syst Bot* **24:** 32–57.

Castellano, M.A., Beever, R.E., and Trappe, J.M. (2012) Sequestrate fungi of New Zealand: *Elaphomyces* and *Rupticutis* gen. nov. (Ascomycota, Eurotiales, Elaphomycetaceae). *N Z J Bot* **50:** 423–433.

Cissé, O.H., Pagni, M., and Hauser, P.M. (2013) De novo assembly of the *Pneumocystis jirovecii* genome from a single bronchoalveolar lavage fluid specimen from a patient. *mBio* **4:** 1–4.

Copeland, A., Lapidus, A., Del Rio, T.G., Nolan, M., Lucas, S., Chen, F., *et al*. (2009) Complete genome sequence of *Catenulispora acidiphila* type strain (ID 139908). *Stand Genomic Sci* **1:** 119–125.

Crawford, D.L., Lynch, J.M., Whipps, J.M., and Ousley, M.A. (1993) Isolation and characterization of Actinomycete antagonists of a fungal root pathogen. *Appl Environ Microbiol* **59:** 3899–3905.

Darling, A.E., Mau, B., and Perna, N.T. (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **5:** e11147.

Desirò, A., Salvioli, A., Ngonkeu, E.L., Mondo, S.J., Epis, S., Faccio, A., *et al*. (2014) Detection of a novel intracellular microbiome hosted in arbuscular mycorrhizal fungi. *ISME J* **8:** 257–270.

Desjardins, C.A., Champion, M.D., Holder, J.W., Muszewska, A., Goldberg, J., Bailão, A.M., *et al*. (2011) Comparative genomic analysis of human fungal pathogens causing paracoccidioidomycosis. *PLoS Genet* **10:** e1002345.

Deveau, A., Palin, B., Delaruelle, C., Peter, M., Kohler, A., Pierrat, J.C., *et al*. (2007) The Mycorrhiza Helper *Pseudomonas Fluorescens* BBc6R8 has a specific priming effect on the growth, morphology and gene expression of the ectomycorrhizal fungus *Laccaria bicolor* S238N. *New Phytol* **175:** 743–755.

Dixon, P. (2003) VEGAN, a package of R functions for community ecology. *J Veg Sci* **14:** 927–930.

Duponnois, R., and Kisa, M. (2006) The possible role of trehalose in the mycorrhiza helper bacterium effect. *Can J Bot* **84:** 1005–1008.

Eddy, S.R. (2011) Accelerated profile HMM searches. *PLoS Comput Biol* **7:** e1002195.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32:** 1792–1797.

Elbein, A.D., Pan, Y.T., Pastuszak, I., and Carroll, D. (2003) New insights on trehalose: a multifunctional molecule. *Glycobiology* **13:** 17R–27R.

Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30:** 1575–1584.

Fedorova, N.D., Khaldi, N., Joardar, V.S., Maiti, R., Amedeo, P., Anderson, M.J., *et al*. (2008) Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. *PLoS Genet* **4:** e1000046.

Fox, E.M., and Howlett, B.J. (2008) Secondary metabolism: regulation and role in fungal biology. *Curr Opin Microbiol* **11:** 481–487.

Frey-Klett, P., and Garbaye, J. (2005) Mycorrhiza helper bacteria: a promising model for the genomic analysis of fungal-bacterial interactions. *New Phytol* **175:** 743–755.

Frey-Klett, P., Garbaye, J., and Tarkka, M. (2007) The mycorrhiza helper bacteria revisited. *New Phytol* **176:** 22–36.

Frey-Klett, P., Burlinson, P., Deveau, A., Barret, M., Tarkka, M., and Sarniguet, A. (2011) Bacterial-fungal interactions: hyphens between agricultural, clinical, environmental, and food microbiologists. *Microbiol Mol Biol Rev* **75:** 583–609.

Geiser, D.M., Gueidan, C., Miadlikowska, J., Lutzoni, F., Kauff, F., Hofstetter, V., *et al*. (2006) Eurotiomycetes: eurotiomycetidae and chaetothyriomycetidae. *Mycologia* **98:** 1053–1064.

Gordon, A. (2011) *'FASTX-Toolkit' FASTQ/A short-reads pre-processing tools*. URL http://hannonlab.cshl.edu/fastx_toolkit/.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., *et al*. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29:** 644–652.

Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., *et al*. (2014) MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res* **42:** D699–D704.

Gryndler, M., Soukupová, L., Hršelová, H., Gryndlerová, H., Borovička, J., Streiblová, E., and Jansa, J. (2013) A quest for indigenous truffle helper prokaryotes. *Environ Microbiol Rep* **5:** 346–352.

Hawker, L.E., Fraymouth, J., and de la Torre, M. (1967) The identity of Elaphomyces granulatus. *Trans Br Mycol Soc* **50:** 129–136.

Hayat, R., Ali, S., Amara, U., Khalid, R., and Ahmed, I. (2010) Soil beneficial bacteria and their role in plant growth promotion: a review. *Ann Microbiol* **60:** 579–598.

Hibbett, D.S., and Matheny, P.B. (2009) The relative ages of ectomycorrhizal mushrooms and their plant hosts estimated using Bayesian relaxed molecular clock analyses. *BMC Biol* **7:** 13.

Hibbett, D.S., and Thorn, R.G. (2001) Basidiomycota: homobasidiomycetes. In *The Mycota VII B: Systematics and Evolution*. McLaughlin, D.J., McLaughlin, E.G., and Lemke, P.A. (eds). Berlin, Germany: Springer, pp. 121–168.

Hoffman, M.T., Gunatilaka, M.K., Wijeratne, K., Gunatilaka, L., and Arnold, A.E. (2013) Endohyphal bacterium enhances production of indole-3-acetic acid by a foliar fungal endophyte. *PLoS ONE* **8:** e73132.

Hosaka, K., Bates, S.T., Beever, R.E., Castellano, M.A., Colgan, W., III, Domínguez, L.S., *et al*. (2006) Molecular phylogenetics of the gomphoid-phalloid fungi with an establishment of the new subclass Phallomycetidae and two new orders. *Mycologia* **98:** 949–959.

Houbraken, J., and Samson, R.A. (2011) Phylogeny of *Penicillium* and the segregation of Trichocomaceae into three families. *Stud Mycol* **70:** 1–51.

Inderbitzin, P., Lim, S.R., Volkmann-Kohlmeyer, B., Kohlmeyer, J., and Berbee, M.L. (2004) The phylogenetic position of *Spathulospora* based on DNA sequences from dried herbarium material. *Mycol Res* **108:** 737–748.

Inglis, D.O., Binkley, J., Skrzypek, M.S., Arnaud, M.B., Cerqueira, G.C., Shah, P., *et al*. (2013) Comprehensive annotation of secondary metabolite biosynthetic genes and gene clusters of *Aspergillus nidulans*, *A. fumigatus*, *A. niger*, and *A. oryzae*. *BMC Microbiol* **12:** 91.

Keller, N.P., Turner, G., and Bennett, J.W. (2005) Fungal secondary metabolism – from biochemistry to genomics. *Nat Rev Microbiol* **3:** 937–947.

Khaldi, N., Seifuddin, F.T., Turner, G., Haft, D., Nierman, W.C., Wolfe, K.H., and Fedorova, N.D. (2010) SMURF: genomic mapping of fungal secondary metabolite clusters. *Fungal Genet Biol* **47:** 736–741.

Kirk, P.M., Cannon, P.F., Minter, D.W., and Stalpers, J.A. (eds) (2008) *The Dictionary of the Fungi*, 10th edn. Wallingford, UK: CAB International.

Korf, R.P. (1973) Discomycetes and tuberales. *Fungi* **IV A:** 249–319.

Kurth, F., Zeitler, K., Feldhahn, L., Neu, T.R., Weber, T., Krištůfek, V., *et al*. (2013) Detection and quantification of a mycorrhization helper bacterium and a mycorrhizal fungus in plant-soil microcosms at different levels of complexity. *BMC Microbiol* **13:** 205.

Landvik, S., Shaller, N.F.J., and Eriksson, O.E. (1996) SSU rDNA sequence support for a close relationship between the Elaphomycetales and the Eurotiales and Onygenales. *Mycoscience* **37:** 237–241.

Langmead, B., and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359.

Lehr, N.A., Schrey, S.D., Bauer, R., Hampp, R., and Tarkka, M.T. (2007) Suppression of plant defense response by a mycorrhiza helper bacterium. *New Phytol* **174:** 892–903.

Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., *et al*. (2010) The sequence and de novo assembly of the giant panda genome. *Nature* **463:** 311–317.

López, M.F., Männer, P., Willmann, A., Hampp, R., and Nehls, U. (2007) Increased trehalose biosynthesis in Hartig net hyphae of ectomycorrhizas. *New Phytol* **174:** 389–398.

LoBuglio, K.F., Berbee, M.L., and Taylor, J.W. (1996) Phylogenetic origins of the asexual mycorrhizal symbiont *Cenococcum geophilum* Fr. and other mycorrhizal fungi among the Ascomycetes. *Mol Phylogenet Evol* **6:** 287–294.

Marçais, G., and Kingsford, C. (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27:** 764–770.

Martin, F., Aerts, A., Ahren, D., Brun, A., Danchin, E.G.J., Duchaussoy, F., *et al*. (2008) The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **452:** 88–93.

Martin, F., Kohler, A., Murat, C., Balestrini, R., Coutinho, P.M., Jaillon, O., *et al*. (2010) Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* **464:** 1033–1038.

Maser, C., Trappe, J.M., and Nussbaum, R.A. (1978) Fungal-small mammal interrelationships with emphasis on Oregon coniferous forests. *Ecology* **59:** 799–809.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., *et al*. (2008) The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9:** 386.

Miller, O.K., Jr, Henkel, T.W., James, T.Y., and Miller, S.L. (2001) *Pseudotulostoma*, a remarkable new volvate genus in the Elaphomycetaceae from Guyana. *Mycol Res* **105:** 1268–1272.

Miller, S.L., and Miller, O.K., Jr (1984) *Synthesis of* Elaphomyces muricatus + Pinus sylvestris *ectomycorrhizae*. *Can J Bot* **62:** 2363–2369.

Mondo, S.J., Toomer, K.H., Morton, J.B., Lekberg, Y., and Pawlowska, T.E. (2012) Evolutionary stability in a 400-million-year-old heritable facultative mutualism. *Evolution* **66:** 2564–2576.

Morgenstern, I., Powlowski, J., Ishmael, N., Darmond, C., Marqueteau, S., Moisan, M.-C., *et al*. (2012) A molecular phylogeny of thermophilic fungi. *Fungal Biol* **116:** 489–502.

Moulin, L., Munive, A., Dreyfus, B., and Boivin-Masson, C. (2001) Nodulation of legumes by members of the β-subclass of Proteobacteria. *Nature* **411:** 948–950.

Moulin, L., Agnieszka, A., Caroline, B., Booth, K., Vriezen, J.A., Melkonian, R., and Riley, M. (2014) Complete Genome sequence of *Burkholderia phymatum* STM815, a broad host range and efficient nitrogen-fixing symbiont of *Mimosa* species. *Stand Genomic Sci* **9:** 763–774.

Nierman, W.C., Pain, A., Anderson, M.J., Wortman, J.R., Kim, H.S., Arroyo, J., *et al*. (2005) Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* **438:** 1151–1156.

Offre, P., Pivato, B., Mazurier, S., Siblot, S., Berta, G., Lemanceau, P., and Mougel, C. (2008) Microdiversity of

Burkholderiales associated with mycorrhizal and nonmycorrhizal roots of *Medicago truncatula*. *FEMS Microbiol Ecol* **65:** 180–192.

O'Hanlon, K.A., Gallagher, L., Schrettl, M., Jöchl, C., Kavanagh, K., Larsen, T.O., and Doyle, S. (2012) Nonribosomal peptide synthetase genes pesL and pes1 are essential for fumigaclavine C production in *Aspergillus fumigatus*. *Appl Environ Microbiol* **78:** 3166–3176.

Parke, J.L., and Gurian-Sherman, D. (2001) Diversity of the *Burkholderia cepacia* complex and implications for risk assessment of biological control strains. *Annu Rev Phytopathol* **39:** 225–258.

Parra, G., Bradnam, K., and Korf, I. (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23:** 1061–1067.

Parra, G., Bradnam, K., Ning, Z., Keane, T., and Korf, I. (2009) Assessing the gene space in draft genomes. *Nucleic Acids Res* **37:** 289–297.

Partida-Martinez, L.P., Groth, I., Schmitt, I., Richter, W., Roth, M., and Hertweck, C. (2007) *Burkholdia rhizoxinica* sp. nov. and *Burkholderia endofungorum* sp. nov., bacterial endosymbionts of the plant-pathogenic fungus *Rhizopus microsporus*. *Int J Syst Evol Microbiol* **57:** 2583–2590.

Patel, S.S., and Donmez, I. (2006) Mechanisms of helicases. *J Biol Chem* **281:** 18265–18268.

Pel, H.J., De Winde, J.H., Archer, D.B., Dyer, P.S., Hofmann, G., Schaap, P.J., *et al*. (2007) Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. *Nat Biotechnol* **25:** 221–231.

Price, A.L., Jones, N.C., and Pevzner, P.A. (2005) *De novo* identification of repeat families in large genomes. *Bioinformatics* **21:** i351–i358.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., *et al*. (2012) The Pfam protein families database. *Nucleic Acids Res* **40:** D290–D301.

Reeves, E.P., Reiber, K., Neville, C., Scheibner, O., Kavanagh, K., and Doyle, S. (2006) A nonribosomal peptide synthetase (Pes1) confers protection against oxidative stress in *Aspergillus fumigatus*. *FEBS J* **273:** 3038–3053.

Renesto, P., Crapoulet, N., Ogata, H., La Scola, B., Vestris, G., Claverie, J.M., and Raoult, D. (2003) Genome-based design of a cell-free culture medium for *Tropheryma whipplei*. *Lancet* **362:** 447–449.

Riedlinger, J., Schrey, S.D., Tarkka, M.T., Hampp, R., Kapur, M., and Fiedler, H.-P. (2006) Auxofuran, a novel metabolite that stimulates the growth of Fly Agaric, is produced by the mycorrhiza helper bacterium *Streptomyces* strain AcH 505. *Appl Environ Microbiol* **72:** 3550–3557.

Robbertse, B., Yoder, R.J., Boyd, A., Reeves, J., and Spatafora, J.W. (2011) Hal: an automated pipeline for phylogenetic analyses of genomic data. *PLoS Curr* **3:** RRN1213.

Samson, R.A., Yilmaz, N., Houbraken, J., Spierenburg, H., Seifert, K.A., Peterson, S.W., *et al*. (2011) Phylogeny and nomenclature of the genus *Talaromyces* and taxa accommodated in *Penicillium* subgenus *Biverticillium*. *Stud Mycol* **70:** 159–183.

Sharpton, T.J., Stajich, J.E., Rounsley, S.D., Gardner, M.J., Wortman, J.R., Jordar, V.S., *et al*. (2009) Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives. *Genome Res* **19:** 1722–1731.

Smit, A.F., and Green, P. (1996) *RepeatMasker*. Published on the web at http://www.Repeatmasker.org.

Spanu, P.D., Abbott, J.C., Amselem, J., Burgis, T.A., Soanes, D.M., Stüber, K., *et al*. (2010) Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* **330:** 1543–1546.

Splivallo, R., Deveau, A., Valdez, N., Kirchhoff, N., Frey-Klett, P., and Karlovsky, P. (2014) Bacteria associated with truffle-fruiting bodies contribute to truffle aroma. *Environ Microbiol* **17:** 2647–2660.

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688–2690.

Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7:** 62.

Talavera, G., and Castresana, J. (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56:** 564–577.

Tedersoo, L., Hansen, K., Perry, B.A., and Kjøller, R. (2006) Molecular and morphological diversity of pezizalean ectomycorrhiza. *New Phytol* **170:** 581–596.

Tedersoo, L., May, T.W., and Smith, M.E. (2010) Ectomycorrhizal lifestyle in fungi: global diversity, distribution, and evolution of phylogenetic lineages. *Myocorrhiza* **20:** 217–263.

Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y.O., and Borodovsky, M. (2008) Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. *Genome Res* **18:** 1979–1990.

Thiers, H.D. (1984) The secotioid syndrome. *Mycologia* **76:** 1–8.

Traeger, S., Altegoer, F., Freitag, M., Gabaldon, T., Kempken, F., Kumar, A., *et al*. (2013) The genome and development-dependent transcriptomes of *Pyronema confluens*: a window into fungal evolution. *PLoS Genet* **9:** e1003820.

Trappe, J.M. (1979) The orders, families, and genera of hypogeous Ascomycotina (truffles and their relatives). *Mycotaxon* **9:** 297–340.

Ward, N.L., Challacombe, J.F., Janssen, P.H., Henrissat, B., Coutinho, P.M., Wu, M., *et al*. (2009) Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. *Appl Environ Microbiol* **75:** 2046–2056.

Wiemken, V. (2007) Trehalose synthesis in ectomycorrhizas – a driving force of carbon gain for fungi? *New Phytol* **174:** 228–230.

Wolfe, B.E., Tulloss, R.E., and Pringle, A. (2012) The irreversible loss of a decomposition pathway marks the single origin of an ectomycorrhizal symbiosis. *PLoS ONE* **7:** e39597.

Wrighton, K.C., Thomas, B.C., Sharon, I., Miller, C.S., Castelle, C.J., VerBerkmoes, N.C., *et al*. (2012) Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* **337:** 1661–1665.

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

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Workflow. Graphical representation of the workflow for obtaining the contigs of the target genome, *E. granulatus*.
Fig. S2. Mitochondrial genome alignment. Progressive Mauve alignment of mitochondrial genome assemblies of *E. granulatus* and *Aspergillus fumigatus*. Coloured blocks represent homologous sequence free of rearrangements, and height within blocks correspond to sequence conservation.
Fig. S3. Adenylation domain phylogeny. Red box encloses the Adenylation domains specific to an *E. granulatus* NRPS that are part of a species-specific lineage expansion. Species abbreviations are as follows: Aspfu, *Aspergillus fumigatus*; Anid, *Aspergillus nidulans*; B_bassiana, *Beauveria bassiana*; Ch, *Cochliobolus heterostrophus*; Cp, *Claviceps purpurea*; Ef, *Epichloë festucae*; Egran, *E. granulatus*; Fe, *Fusarium equiseti*; Fh, *Fusarium heterosporum*; Hv, *Trichoderma virens*; Lm, *Leptosphaeria maculans*; Ma/MAA, *Metarhizium anisopliae*; Mg, *Magnaporthe grisea*; Pench, *Penicillium chrysogenum*; Talstip, *Talaromyces stipitatus*; TOPH, *Tolypocladium ophioglossoides* (parasite of *Elaphomyces* spp.).
Fig. S4. Species richness rarefaction. Rarefaction curves for species richness versus sampling depth for the two peridium libraries with paired-end reads uploaded separately (2 reads per library). MG-RAST IDs are given to the right of the curves.
Fig. S5. Taxonomy Pie Chart. Percentages of the microbiome reads in the *E. granulatus* peridium annotated at the domain level.
Fig. S6. Rarefaction of ECs. Rarefaction of the enzyme commission numbers for *E. granulatus* peridium microbiome with subsample size of 250.
Fig. S7. KEGG Map. KeggMapper visualization of the Enzyme Categories present in the *E. granulatus* peridium annotated by MG-RAST (highlighted in purple).
Table S1. List of 35 bacteria. List of the 35 bacterial genomes used in the bacterial database to removed raw reads using Bowtie 2.
Table S2. Genome assembly statistics for *Caliciopsis orientalis* and *Monascus ruber*, both of which are being published for the first time here. *Sequenced and assembled at Oregon State University using methods described in text. **Sequenced and assembled at the Joint Genomes Institute as a part of 1KFG CSP. ***Ab initio* models created in AUGUSTUS – no RNA sequenced.
Table S3. Table of CAZymes. Counts of CAZymes for the selected Eurotiomycetes and Pezizomycetes analysed in this study, including *E. granulatus* (highlighted in yellow), other Eurotiales (green), Onygenales (pink), *Caliciopsis orientalis* (olive) and Pezizales (blue). Species abbreviations as in Fig. 3. CAZyme class/module abbreviations: auxiliary activities (AA), carbohydrate-binding module (CBM), carbohydrate esterase (CE), glycoside hydrolase (GH), glycosyltransferase (GT), polysaccharide Lyase (PL).
Table S4. Overrepresented clusters in *E. granulatus*. MCL clusters, generated through the Hal pipeline at 1.2 inflation parameter, in which *E. granulatus* is overrepresented compared with species in Fig. 2. To conserve space, non-Eurotiomycetes have been removed, except for *Tuber melanosporum* (Tum). The average number of protein models per cluster per taxon for all taxa in Fig. 2 is presented in the last column. Putative annotations are given, where known. Species abbreviations as in Fig. 3.