The genome and microbiome of a dikaryotic fungus (*Inocybe terrigena*, Inocybaceae) revealed by metagenomics

Mohammad Bahram[1,2*], Dan Vanderpool[3], Mari Pent[2], Markus Hiltunen[1], Martin Ryberg[1]

[1]*Department of Organismal Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 75236 Uppsala, Sweden* [2]*Department of Botany, Institute of Ecology and Earth Sciences, University of Tartu, 40 Lai St., 51005 Tartu, Estonia* [3]*Division of Biological Sciences, University of Montana, 32 Campus Drive, Missoula, MT 59812, USA*

Abstract

Recent advances in molecular methods have increased our understanding of various fungal symbioses. However, little is known about genomic and microbiome features of most uncultured symbiotic fungal clades. Here, we analysed the genome and microbiome of Inocybaceae (Agaricales, Basidiomycota), a largely uncultured ectomycorrhizal clade known to form symbiotic associations with a wide variety of plant species. We used metagenomic sequencing and assembly of dikaryotic fruiting-body tissues from *Inocybe terrigena* (Fr.) Kuyper, to classify fungal and bacterial genomic sequences, and obtained a nearly complete fungal genome containing 93% of core eukaryotic genes. Comparative genomics reveals that *I. terrigena* is more similar to previously published ectmycorrhizal and brown rot fungi than white rot fungi. The reduction in lignin degradation capacity has been independent from, and significantly faster than in, closely related ectomycorrhizal clades suggesting that ectomycorrhizal symbiosis evolved independently in *Inocybe*. The microbiome of *I. terrigena* fruiting-bodies includes bacteria with known symbiotic functions in other fungal and non-fungal host environments, suggesting potential symbiotic functions of these bacteria in fungal tissues regardless of habitat conditions. Our study demonstrates the usefulness of direct metagenomics analysis of fruiting-body tissues for characterizing fungal genomes and microbiome.

*: *Corresponding author: Mohammad Bahram, E-mail address: bahram@ut.ee*

Introduction

Ectomycorrhizal fungi constitute a major component of fungal communities in terrestrial ecosystems, functioning to facilitate nutrient uptake and carbon storage by plants (Smith and Read 2010). Evidence from marker-gene studies, particularly ribosomal RNA genes, suggests that ectomycorrhizal fungi have evolved independently in several fungal lineages (Hibbett *et al*. 2000; Tedersoo *et al*. 2010; Veldre *et al*. 2013) but the genomic adaptations associated with these functional shifts are only beginning to be characterized (Kohler *et al*. 2015).

Recent genomic studies focusing on fungi and using high-throughput sequencing (HTS) technologies have improved our understanding of the evolution of mycorrhizal symbiosis. These studies provide evidence that while mycorrhizal fungi have retained much of their enzymatic capacity to release nutrients from complex organic compounds, certain carbohydrate active enzyme (CAZyme) families have contracted in the genome evolution of mycorrhizal fungi compared to their saprotrophic ancestors (Martin *et al*. 2010; Kohler *et al*. 2015). In addition, metabolite pathways and secondary metabolites vary between obligate biotrophs and saprotrophs reflecting evolutionary adaptation to a shift in life history, contributing to variation in genome size, characteristic reductions in gene families such as transporters and plant cell wall degradation enzymes (Martin *et al*. 2010; Spanu *et al*. 2010; Floudas *et al*. 2015; Kohler *et al*. 2015). Ectomycorrhizal fungi possess a repertoire of genes encoding cellulose degrading enzymes to release simple organic compounds that are available for uptake by plants, which substantially accelerate soil nutrient cycling (Martin *et al*. 2010). Nevertheless, most fungal genome sequencing studies have employed culture-dependent techniques, which omit major ectomycorrhizal lineages that are difficult to culture.

Similar to animals and plants, fungal tissues can harbour a diverse array of prokaryote associates. In root and soil fungi, bacteria may contribute to the formation and regulation of mycorrhizal associations (Torres-Cortés *et al*. 2015). A few studies on fungal fruiting-bodies suggest that bacteria may have important ecological roles in fungal spore dispersal (Splivallo *et al*. 2015), gene expression (Riedlinger *et al*. 2006; Deveau *et al*. 2007) and mycotoxin production (Lackner *et al*. 2009). In addition, bacteria may affect fungal growth (Chen *et al*.

2013) and mycorrhization (Frey-Klett *et al*. 2007; Aspray *et al*. 2013), yet we know little about the associated bacterial taxa and functions in epigeous fruiting-bodies.

We performed metagenomic sequencing on DNA extracted from fruiting-body tissues to characterize the genomes of the dikaryotic fungus *Inocybe terrigena* (Inocybaceae, Agaricales) and those of its associated bacterial community. Inocybaceae is a diverse ectomycorrhizal fungal lineage that is thought to have evolved independently from other ectomycorrhizal lineages (Matheny *et al*. 2009). Nevertheless, there is no published work on the evolution of mycorrhizal status in Inocybaceae using comparative genomics. We compared protein functional domains (PFAM) and CAZymes for the genomes of *I. terrigena* and its closely related clades to look for significant expansion or contraction in gene families across multiple, independent evolutionary transitions to mycorrhizal symbiosis (Matheny *et al*. 2009). Additionally, we characterized the taxonomic composition and potential functions of the associated bacteria in fruiting-body tissues.

Material and Methods

DNA was extracted from a dried collection of *Inocybe terrigena* (field accession no: MR00339; coordinates: N59°0'7.52, E17°42'2.42"; date: 2013-09-20; herbarium: UPS), collected and identified by MR, using Plant mix DNeasy DNA Isolation kit. To obtain DNA of sufficient quantity, four DNA extractions from lamellae of the fruiting-body were pooled. Lamellae are internal spore bearing layered structures in fruiting-bodies and are therefore expected to provide higher DNA per unit of material. In addition, as lamellae are protected during development they have a lower risk for contamination than other exposed parts of the fruiting-body. PCR free, paired-end, 300bp insert libraries were constructed from total DNA and sequenced on 1/10 of an Illumina HiSeq 2500 lane at Sci-Life laboratory (Stockholm, Sweden). Raw Illumina reads, available at the Sequence Read Archive (SRA) under accession number SRP066410, were quality filtered and trimmed using Fastx Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/[2]). Genome assembly was performed using Spades de novo (Bankevich *et al*. 2012) with default settings. To sort assembled contigs into fungi and bacteria, we used BLAST searchesto compare contigs to previously published whole genomes of bacteria and fungi downloaded from GenBank and the Joint Genome Institute (JGI; www.jgi.doe.gov) website, respectively. To evaluate the accuracy and coverage of the resulting assemblies, we mapped post-filtered reads to the

assembled contigs using Bowtie (Langmead *et al*. 2012). We used median coverage and GC content to determine whether contigs not classified by BLAST were of bacterial or fungal origin, by comparison to the values of contigs classified by BLAST. Unclassified contigs shorter than 1000 bp were excluded from the analysis, as GC content and median coverage varied greatly for these and taxonomic assignment was deemed unreliable. Bacterial and eukaryotic species present in the final assemblies were identified using (previously constructed) HMMer profiles to recognize and trim ribosomal operons and ITS region, respectively (Lagesen *et al*. 2007; Bengtsson-Palme *et al*. 2013). Additionally, we used the MG-RAST pipeline (Meyer *et al*. 2008) to infer both taxonomic and functional annotations based on assemblies.

Protein annotation of fungal and bacterial assemblies was performed using the Maker2.3.36 pipeline (Holt & Yandell 2011) respectively Prodigal (Hyatt *et al*. 2010) with default parameters. Core Eukaryotic Mapping Genes Approach (CEGMA) (Parra *et al*. 2007) was used to evaluate genome completeness and to generate preliminary annotations as training sets in maker. RepeatModeler (Smit & Hubley 2011) was used to generate a classified repeat library for the metagenome assembly. This repetitive sequence library was combined with a Maker2 provided transposable element library and was used in RepeatMasker 3.0 (Smit *et al*. 2010) for masking within the Maker2 pipeline. Additionally, the proteomes of *Laccaria_bicolor*, *Coprinopsis_cinerea* and the UniProt reference proteomes (The UniProt Consortium, 2014) were used as protein evidence to generate hints for *ab-initio* predictors. Three rounds of training were used to for the *ab-initio* gene predictors SNAP and Augustus while self-training was used for GeneMark-ES. Overcalling genes is common for *ab-initio* gene predictors (Larsen & Krogh 2003) so "keep_preds=0" was set in Maker2 to only call gene models which protein evidence (AED<1) supporting them. Predicted proteins lacking functional evidence were scanned for protein family domains (PFAM) using InterProScan5 (IPRscan5) (Jones *et al*. 2014), those containing PFAM domains were promoted to full gene models using scripts included with Maker2. Protein models were assigned to multigene families using the OrthoMCL software (v2.0.9) (Li *et al*. 2003). BLASTX was used to compare and sort bacterial and fungal protein models against the whole NCBI protein database. Analyses were facilitated with a custom perl pipeline, which is available in the supplementary information. Fungal protein models were compared to non-redundant genomes using BlastKOALA (Kanehisa *et al*. 2016) for functional

annotation with additional screening for carbohydrate-active enzymes (CAZYmes) using the dbCAN pipeline (Yin *et al*. 2012).

Using the mapped read dataset from Bowtie, duplicate reads were marked with Picard Tools and reads were realigned around indels with Genome Analysis Toolkit (McKenna *et al*. 2010). Variant calling was performed on the realigned reads using Platypus (Rimmer *et al*. 2014). The level of heterozygosity was quantified from the filtered variant dataset.

Phylogenetic analysis was performed based on complementary supermatrix and supertree approaches with 74 publicly available Agaricomycetes genomes (retrieved from the JGI database). Single-copy genes were identified based on MCL clusters on BLAST e-values with the heath parameter set to 2.0, in an additional quality check they were clustered using OrthoMCL with only those single copy genes present in 75% of taxa used for phylogenetic analysis. Amino acid sequences were aligned using MAFFT 7 (Katoh & Standley 2013) with highly variable regions removed using Gblocks (Talavera *et al*. 2007). The resulting concatenated alignment was used to estimate a maximum likelihood phylogeny with 1,000 UltraFast (UF) bootstrap replicates (Minh *et al*. 2013) using IQ-Tree 1.5.5 (Nguyen *et al*. 2015). Sequence evolution was modeled using the Posterior Mean Site Frequency (PSMF) model (Wang *et al*. 2017) and 60 mixture classes. PhyloBayes 3.3 (Lartillot *et al*. 2009) with site-specific evolutionary rates modeled using non-parametric infinite mixtures (CAT-GTR) was used to generate a fossil calibrated ultrametric tree from our whole genome concatenated tree. A diffuse gamma prior with the mean equal to the standard deviation was used for the root of the tree. A log-normal auto-correlated relaxed clock model (Thorne *et al*. 1998) with a Dirichlet prior on divergence times was used for dating the tree. Two fossils were employed to calibrate node times to geological time. The first corresponds to a minimum age of 90My for the Agaricales (Hibbett *et al*. 1997). The second corresponds to a minimum age of 360My (Stubblefield *et al*. 1985) for the Basidiomycota in general. Trees for individual genes were constructed using the PROTGAMMAWAG model and 100 bootstraps in RAxML 8.2.4 (Stamatakis 2014). The trees were then combined to species trees using ASTRAL; support values were calculated using the bootstrap trees (Mirarab & Warnow 2015). Significant expansion or contraction of protein family domains (PFAM) was determined for a reduced phylogeny of the 59 taxa belonging to Agaricomycotina using CAFE 3.1 (Han *et al*. 2013).

PFAM domains found to be significantly expanded or contracted were characterized using dcGO (Fang & Gough 2012).

Results

Genomic features of Inocybe terrigena

After quality filtering and trimming 29,244,774 paired end reads (96%) were used for assembly. Statistics for sequencing and assembly are available in Table S1. BLAST searchers performed with the resulting 31,471 assembled contigs against reference genomes, resulted in 2,262 identified as fungal (N50: 30,623, total length: 26,127,495 bp) and 11,596 as bacterial (N50: 8,777, total length 34,853,629 bp). There were 417 fungal and 17,196 bacterial contigs that were uncharacterized by our BLAST searches. The MGRAST analysis revealed that 77.42% of contigs were of bacterial origin followed by 22.47% fungal, 0.04% viral and 0.03% archaeal contigs. Mapping all reads to the assembly resulted in a total of 28,632,347 mapped reads (93% of all reads), with a median coverage of 35X in the fungal contigs (22,274,636 reads; 78% of mapped reads) and 5X for bacterial contigs (5,072,696 reads; 18%). Variant calling revealed 3.4 variant sites per kb in the fungal contigs.

BLAST analyses of extracted ITS sequences against UNITE (Kõljalg *et al*. 2013) resulted in one ITS sequence matching *Inocybe terrigena* (accession number: JF908091; identity=99%, e-value=1e-137, median coverage=2335, contig length=8222), and five additional ITS sequences matching other Eukaryotes. All of the additional eukaryotic contigs including ITS had very low coverage: Hypomyces odoratus (identity=100%, e-value=1e-86, median coverage=2, contig length=745), Selaginella deflexa (identity=98%, e-value=6e-102, median coverage=2, contig length=1175), Drosophila subauraria (identity=84%, e-value=3e-23, median coverage=1, contig length=438), Cucumis melo (identity=100%, e-value=3e-04, median coverage=3, contig length=1041). These results indicate a negligible contribution of other Eukaryotes to the resulting *I. terrigena* assembly.

The Maker2 gene annotation of fungal (n=2262), unclassified (n=17196) and ambivalent (n=417) contigs resulted in 11,918 fungal genes identified (coverage=50.58±142.49, median±SD; GC content=0.467±0.0295). The Prodigal annotation of bacterial contigs resulted

in 63,328 bacterial genes. Of these, 3,289 additional fungal genes were identified based on Blast searches (median coverage=57.33±152.85; GC content=0.471±0.043), indicating some fungal contigs were misidentified as bacteria. Thus, we confirmed the identification of all classified contigs based on BLAST searches using annotated genes against the NCBI protein database, as implemented in the pipeline (Table S4; Supporting Information 1). Based on genes annotated by Prodigal and genes identified from unclassified contigs using Maker2 (n=8190, median coverage= 10.70±11.77; GC content=0.599±0.065), 53,271 genes were identified as bacterial (median coverage=10.07±13.24; GC content=0.588±0.080). After this additional round of filtering, 2,855 genes remain unclassified (median coverage=6.36±37.47; GC content=0.598±0.073).

In total, the *I. terrigena* genome included 15,207 genes, which is lower than those reported for most ectomycorrhizal fungi (Fig. S1; Table S5). Of 15,207 genes, 24.6% genes were with Pfam domains based on InterPro domain assignments. Using KEGGmapper, 4,681 out of 15,207 fungal genes were functionally annotated (30.8%; Fig. S2; Table S2). Using CAZY pipeline, we identified 985 Pfam domains from CAZy Families, including 75 auxiliary activities (AA, 11 families), 75 carbohydrate-binding module (CBM, 10 families), 143 carbohydrate esterases (CE, 8 families), 334 glycoside hydrolase (GH, 43 families) and 371 glycosyltransferases (GT, 47 families) domains. Similar to the genomes of other related ectomycorrhizal species, we found no lignin-modifying fungal peroxidases (PODs) in the *I. terrigena* genome.

The CEGMA analysis indicated 91% of 242 full-length, core eukaryotic genes were present in our fungal assembly and 93% were present when partial and full-length alignments were considered. (Table S3). Longest and shortest fungal scaffolds were 241,864 bp and 170 bp respectively.

Mitochondrial contigs were determined based on coverage (>1200X) and GC content (<0.40), resulting in 5 contigs, with a combined length of 67 kb. 85.8% of non mitochondrial fungal sequences had a GC content between 0.4 and 0.5 (Table S4). Most bacterial contigs are outside this interval (97.2%). Among non-fungal genera, *Pedobacter*, *Pseudomonas* and *Burkholderia* had the highest coverages of 40.24±20.64, 5.73±3.24 and 4.23±1.99, respectively

(Fig. S3). In total, 56.5% and 10.6% of total bacterial assembly length belonged to *Pseudomonas* and Pedobacter, with GC content of 0.608±0.032 and 0.396±0.039, respectively (Fig. 1; Fig. 8 dominant genus). The remaining contigs of bacterial origin had GC content of 0.500±0.115 and coverage of 18.86±108.85 (Fig. 1; Fig. S3). Based on this we classified previously unclassified contigs with median coverage of >1 and GC content >0.35 and <0.56 (Fig. 1) as fungal, but made no further attempts classifying contigs as bacteria. The final assembly included 5410 fungal contigs (N50: 19,600, total length: 37,436,744 bp) and 12,040 bacterial contigs (N50, 9,606, total length 39,035,878 bp; Table S4).

Comparative genomics

OrthoMCL clustering of 74 Agaricomycetes whole proteomes resulted in 870 single-copy genes present in 75% of taxa, and a concatenated alignment of 148,316 amino acid positions. The resulting Maximum Likelihood phylogenetic tree (Fig. 2) is largely consistent with previous phylogenies (e.g. Kohler *et al*. 2015) and the ASTRAL phylogeny (Supplementary Information 2) in all well supported parts. ATPases and ATP-binding domain of ABC transporters showed the highest expansion and contraction in *I. terrigena* compared to the most closely related mycorrhizal species, *Laccaria bicolor*. Our CAFE analysis of CAZYmes indicated that 6 of ~220 CAZY families found were significantly expanded or contracted in the 59 genomes analysed here (Fig. 3). Cluster analysis revealed two major groups of fungi based on CAZyme profiles, and it placed *I. terrigena* in the same cluster as most other ectomycorrhizal taxa (Fig. ~~2).~~ Compared to other genomes included in our comparative genomics analysis, *I. terrigena* has no significantly expanded CAZyme families; however, it is significantly reduced in an important lignin degrading CAZyme family (AA2), which contains class II lignin-modifying peroxidases. While *I. terrigena* belongs to a clade which has gained 3 AA2 from its parent node, it has lost 7 AA2 compared to other members of this clade.

Microbiome of Inocybe terrigena

Proteobacteria and Bacteroidetes comprised 80.9% and 16.6% of the taxonomic composition in bacterial metagenomics data (Fig. 4A). Gammaproteobacteria (49.9%), Betaproteobacteria (26.7%), Flavobacteria (10.3%), Sphingobacteria (4.6%) and Alphaproteobacteria (3.7%) were the dominant bacterial class (Fig. 4B). *Pseudomonas*, *Chryseobacterium, Herbaspirillum*, *Burkholderia* and *Pedobacter* were the dominant genera (Fig. 4D). In total, the α-diversity of

our metagenome was 63 species based on MG-RAST pipeline, with three dominant species *Pseudomonas fluorescens*, *P. syringae* and *Herbaspirillum seropedicae* contributing 61% to the total bacterial genes.

The bacterial contigs contained 28,288 predicted proteins that mostly matched genes with known functions including *metabolism* (40%), *cellular processes and signalling* (21%) and *information storage and processing* (17%). Similarly, *metabolism and environmental* as well as *genetic information processing* were the most common functional categories (Fig. 5).

Discussion

Genomic features of Inocybe terrigena

Our study is the first to analyse the genome and bacterial communities of a dikaryotic fungus. Using one tenth of an Illumina lane, we obtained a genome with a comparable completeness with those reported in previous studies (Kohler *et al*. 2015; Quandt *et al*. 2015), indicating the acceptable performance of this approach to investigate the genomes of unculturable fungi based on epigeous fruiting-body tissues. Such analysis is facilitated by significant differences in length, GC content and median coverage between bacterial and fungal genomes, which enable classification of assemblies into fungal or bacterial origin. The GC content of the *I. terrigena* genome (0.47) is in the same range as its closely related ecotmycorrhizal fungal species such as *Laccaria bicolor* (0.47) and *L. amethystina* (0.47), *Hypholoma sublateritum* (0.51), *Galerina marginata* (0.48), and *Hebeloma cylindrosporum* (0.48). The GC content of all major bacterial groups except three Bacteroidetes (comprising 5.4% of contigs) were outside this range (Table S3 contigs). Together, the contrasting GC content and coverage can be used to accurately separate bacterial and fungal contigs, which can reduce the computation cost and database biases, particularly when specific genes are targeted.

Our analysis revealed that *I. terrigena* has a smaller genome size based on the number of gene models compared to the previously sequenced ectomycorrhizal Basidiomycetes, e.g. *Laccaria bicolor and Hebeloma cylindrosporum* (Fig. S1, Table S5 external genomes, Martin *et al*. 2010; Kohler *et al*. 2015). Our comparative analysis across a set of available fungal genomes revealed a similar CAZyme profile of *I. terrigena* compared to brown rot and ectomycorrhizal

lineages (Fig. 2). Given the significant reduction in certain CAZyme families as a common pattern in genome evolution of ectomycorrhizal fungi (Kohler *et al*. 2015), the increase in the AA2 family in the lineage leading to the most recent common ancestor to *Inocybe* and its sister clade indicates that this line was not ectomycorrhizal; therefore, the significant reduction in AA2 in the lineage leading to *Inocybe* indicates a separate origin of the ectomycorrhizal nutritional mode in this lineage.

Structure and function of associated bacteria of Inocybe terrigena

Proteobacteria and Bacteroidetes were the most abundant phyla in *I. terrigena,* constituting the largest fraction of bacterial assembly (Fig. S3). Similarly to our study, Proteobacteria has been identified as a very common bacterial group in several ascomycetous (Quandt *et al*. 2015; Benucci & Bonito 2016; Barbieri *et al*. 2005, 2007) as well as basidiomycetous (Kumari *et al*. 2013; Pent *et al*. 2017) fruiting-bodies. There is also evidence that the relative abundance of Proteobacteria is higher in the mycosphere (Warmink *et al*. 2009) and ectomycorhizosphere (Uroz *et al*. 2012; Antony-Babu *et al*. 2013) compared to the surrounding soil, suggesting the tendency of Proteobacteria for colonizing fungus related habitats. A high relative abundance of Bacteroidetes, the second-largest bacterial phylum in *I. terrigena*, is also often found in in ectomycorhizosphere, mycosphere and fruiting-bodies of ascomycetous as well as basidiomycetous fungi (Uroz *et al*. 2012; Antony-Babu *et al*. 2013; Benucci & Bonito 2016; Halsey *et al*. 2016; Pent *et al*. 2017). In particular, the relative abundance of Sphingobacteriia in *I. terrigena* was comparable to that in *Elaphomyces granulatus* (Quandt *et al*. 2015). Similarly to *I. terrigena* fruiting-body Acidobacteria, Actinobacteria, Verrucomicrobia, Firmicutes and Cyanobacteria form a small fraction of the bacterial community in basidiomycetous (Pent *et al*. 2017) as in ascomycetous (Antony-Babu *et al*. 2013) fruiting-bodies, whereas they are highly represented in soil (Eilers *et al*. 2010; Bergmann *et al*. 2011). The observed level of specificity of fungal associated bacteria may be related to close fungal-bacterial interactions or fungal-driven changes in habitat conditions, such as change in pH or nutrient conditions in closely related bulk soil (Danell *et al*. 1993; Nazir *et al*. 2010a,b).

The most abundant bacterial genera associated with *I. terrigena* are known to have symbiotic functions in fungal tissues. Although major bacterial taxa found in *I. terrigena* (Fig. S3) are common in soil (Baldani *et al*. 1986; Janssen 2006), the genus *Pseudomonas* is also one

of the most frequently identified bacterial groups in basidiomycetous fruiting-bodies and ectomycorrhizas (Deveau *et al*. 2007; Kumari *et al*. 2013). While *Pseudomonas* strains are mainly saprotrophic, they are also known to alleviate detrimental effect of pathogens on plant roots and leaves (Haas & Defago 2005) and facilitate mycorrhizal establishment *(*Dominguez *et al*. 2012*)*. Some strains are also plant pathogenic such as *Pseudomonas syringae.* Some strains of *P. fluorescens* can promote the growth of mycelium and ascus opening or other morphological changes in fungi (Citterio *et al* 2001; Cho *et al*. 2003; Deveau *et al*. 2007). Certain *Pseudomonas* species have also been reported as bacterial "fungiphiles" in the mycospheres, pointing to their close relationship with fungi (Warmink *et al*. 2009). Furthermore, *Pedobacter* has been identified as fungus related and tolaasin detoxifying bacteria from Agaricales (Tsukamoto *et al*. 2002), and *Chryseobacterium* has been identified in mycospheres of several fungi (Warmink *et al*. 2009). *Herbaspirillum seropedicae* has been reported as nitrogen fixing root associated bacterium (Baldani *et al*. 1986). Another dominant group *Burkholderia* are also known to have beneficial interactions with fungi improving the formation of mycorrhiza (Aspray *et al*. 2006; Frey-Klett *et al*. 2007) and provide the fungal partner with nutrients in stress conditions (Stopnisek *et al*. 2016). It is possible that these bacteria play similar roles in epigeous fungal fruiting-bodies; however, more replicated studies are needed to understand the specific functions of these bacteria and to exclude the possibility that they passively colonize fungal fruiting-bodies. Taken together, these data indicate the dominance of symbiotic bacteria in fungal epigeous fruiting-bodies.

Despite many similarities, there were some differences between the microbiome of basidiomycetous *I. terrigena* and hypogeous ascomycetous fruiting-bodies (Benucci & Bonito 2016; Barbieri *et al*. 2005, 2007;Quandt *et al*. 2015). Particularly compared to our dataset where Gamma and Betaproteobacteria were the dominant classes, Alphaproteobacteria and Actinobacteria were the dominant classes in ascomycetous fruiting-bodies. This may be explained by the higher relative abundance of Alphaproteobacteria in soil (Janssen 2006; Fierer *et al*. 2012; Pent *et al*. 2017) as well as more intimate association between soil and hypogeous fruiting-bodies, compared to epigeous fruiting-bodies. The differences between bacterial communities of epi- and hypogeous fruiting-bodies may also be explained by different environmental conditions below- and above-ground. In contrast to the microbiome of hypogeous species, which are typically dominated by Bradyrhizobium species (Barbieri *et al*. 2005; 2007;

Antony-Babu *et al*. 2013; Quandt *et al*. 2015), both *Pseudomonas and Pedobacter* dominated the *I. terrigena microbiome*. Several potato-associated *Pseudomonas* species are able to counteract both plant-pathogenic fungi and plant-parasitic nematode (Krechel *et al*. 2002), and some *Pedobacter* are associated with soil or plant-pathogenic nematodes (Tian *et al*. 2011; Baquiran *et al*. 2013). Thus, it is tempting to suggest that the high abundance of *Pseudomonas* and *Pedobacter* may have similar functions in epigeous fruiting-bodies.

Comparing the relative abundance of functional gene categories in *I. terrigena* and non-desert soil microbiome (Fierer *et al*. 2012) reveals a similar functional composition between the two distinct environments. This together with remarkable similarity in their taxonomic composition suggests that soil microbes acts as a major species source for fungal associated bacterial communities (Pent *et al*. 2017). Nonetheless, the high relative abundance of genes functionally related to environmental and genetic information processing (Fig. 5, S1) may facilitate processing a large amount of information from their host environment, to enhance mycorrhizal colonization and reduce the impact of harmful environmental conditions and pathogens (Frey-Klett *et al*. 2007).

Conclusions

This study demonstrates that with appropriate filtering metagenomic sequencing of fungal fruiting-body tissues enables near complete genome sequencing from dikaryotic fruiting-body tissues, with comparable completeness to those from cultured fungal isolates. With further advances in HTS technology, e.g. overcoming length limitation and improving assembly algorithms and hence genome assembly quality, metagenomics will also be useful to study the population genomics of uncultured fungi. In addition, metagenomics enabled us to characterize the associated bacterial taxa and functions in fruiting-body tissues. Certain groups of these bacteria are known to have symbiotic functions with a fungal host.

Acknowledgements

We thank Sandra Lorena Ament-Velásquez and Fabien Burki for advice on bioinformatics. The authors would like to acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure, NGI, and Uppmax for providing assistance in massive parallel

sequencing and computational infrastructure. Funding was provided by Uppsala University (Department of Organismal Biology) and Kaptain Carl Stenholm

References

Antony-Babu S., Deveau A., Nostrand JDV., Zhou J., Le Tacon F., Robin C., Frey-Klett P., Uroz S. 2013. Black truffle-associated bacterial communities during the development and maturation of *Tuber melanosporum* ascocarps and putative functional roles. Environmental Microbiology 16:2831–2847.

Aspray TJ., Jones EE., Davies MW., Shipman M., Bending GD. 2013. Increased hyphal branching and growth of ectomycorrhizal fungus *Lactarius rufus* by the helper bacterium Paenibacillus sp. Mycorrhiza 23: 403–410.

Aspray TJ., Jones EE., Whipps J.M., Bending G.D. 2006. Importance of mycorrhization helper bacteria cell density and metabolite localization for the *Pinus sylvestris-Lactarius rufus* symbiosis. FEMS Microbioloy Ecology 56: 25–33.

Baldani JI., Baldani VLD., Seldin L, Döbereiner J. 1986. Characterization of *Herbaspirillum seropedicae* gen. nov., sp. nov., a root-associated nitrogen-fixing bacterium. International Journal of Systematic Bacteriology 36: 86–93.

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

Baquiran JP., Thater B., Sedky S., De Ley P., Crowley D., Orwin P.M. 2013. Culture-Independent investigation of the microbiome associated with the nematode *Acrobeloides maximus*. PLOS ONE 8: e67425.

Barbieri E., Bertini L., Rossi I., Ceccaroli P., Saltarelli R., Guidi C., Zambonelli A., Stocchi V. 2005. New evidence for bacterial diversity in the ascoma of the ectomycorrhizal fungus *Tuber borchii* Vittad. FEMS Microbiology Letters 247:23–35.

Barbieri E., Guidi C., Bertaux J., Frey-Klett P., Garbaye J., Ceccaroli P., Saltarelli R., Zambonelli A., Stocchi V. 2007. Occurrence and diversity of bacterial communities in Tuber magnatum during truffle maturation. Environmental Microbiology 9: 2234–2246.

Bengtsson-Palme J., Ryberg M., Hartmann M., Branco S., Wang Z., Godhe A., Wit P., Sánchez-García M., Ebersberger I., Sousa F. 2013. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. Methods in Ecology and Evolution 4:914–919.

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

Bergmann GT., Bates ST., Eilers KG., Lauber CL., Caporaso JG., Walters WA., Knight R., Fierer N. 2011. The under-recognized dominance of Verrucomicrobia in soil bacterial communities. Soil Biology Biochemistry 43: 1450–1455.

Chen S., Qiu C., Huang T., Zhou W., Qi Y., Gao Y., Shen J., Qiu L. 2013. Effect of 1-aminocyclopropane-1-carboxylic acid deaminase producing bacteria on the hyphal growth and primordium initation of Agaricus bisporus. Fungal Ecology 6: 110–118.

399 Cho Y.S., Kim J.S., Crowley D.E., Cho B.G. 2003. Growth promotion of the edible fungus
400 *Pleurotus ostreatus* by fluorescent pseudomonads. FEMS Microbiology Letters 218:271–
401 276.
402 Citterio B., Malatesta M., Battistelli S., Marcheggiani F., Baffone W., Saltarelli R., Stocchi V.,
403 Gazzanelli G. 2001. Possible involvement of *Pseudomonas fluorescens* and Bacillaceae in
404 structural modifications of Tuber borchii fruit bodies. Canadian Journal of Microbiology
405 47:264–268.
406 Dahm H., Wrótniak W., Strzelczyk E., Bednarska E. 2005. Diversity of Culturable Bacteria
407 associated with fruiting bodies of ectomycorrhizal fungi. Phytopathologia Polonica 38:51–
408 62.
409 Danell E., Alström S., Ternström A. 1993. *Pseudomonas fluorescens* in association with
410 fruitbodies of the ectomycorrhizal mushroom *Cantharellus cibarius*. Mycological Research
411 97:1148–1152.
412 Deveau A., Palin B., Delaruelle C., Peter M., Kohler A., Pierrat J.C., Sarniguet A., Garbaye J.,
413 Martin F., Frey-Klett P. 2007. The mycorrhiza helper Pseudomonas fluorescens BBc6R8 has
414 a specific priming effect on the growth, morphology and gene expression of the
415 ectomycorrhizal fungus Laccaria bicolor S238N. New Phytologist 175: 743–755.
416 Dominguez JA., Martin A., Anriquez A., Albanesi A. 2012. The combined effects of
417 *Pseudomonas fluorescens* and *Tuber melanosporum* on the quality of *Pinus halepensis*
418 seedlings. Mycorrhiza 22: 429–436.
419 Doré J., Perraud M., Dieryckx C., Kohler A., Morin E., Henrissat B., Lindquist E.,
420 Zimmermann SD., Girard V., Kuo A. 2015. Comparative genomics, proteomics and
421 transcriptomics give new insight into the exoproteome of the basidiomycete Hebeloma
422 cylindrosporum and its involvement in ectomycorrhizal symbiosis. New Phytologist
423 208:1169–1187.
424 Eilers KG., Lauber CL., Knight R., Fierer N. 2010. Shifts in bacterial community structure
425 associated with inputs of low molecular weight carbon compounds to soil. Soil Biology and
426 Biochemistry 42: 896–903.
427 Fang H., Gough J. 2012. dcGO: database of domain-centric ontologies on functions,
428 phenotypes, diseases and more. Nucleic Acids Research 41:D536–D544.
429 Fierer N., Left JW., Adams BJ., Nielsen UN., Bates S.T., Lauber C.L., Owens S., Gilbert J.A.,
430 Wall D.H., Caporaso J.G. 2012. Cross-biome metagenomic analyses of soil microbial
431 communities and their functional attributes. Proceedings of the National Academy of
432 Sciences 109: 21390–21395.
433 Floudas D., Binder M., Riley R., Barry K., Blanchette R.A., Henrissat B., Martinez A.T., *et al*.
434 2012. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal
435 genomes. Science 336: 1715–1719.
436 Frey-Klett P., Garbaye J., Tarkka M. 2007. The mycorrhiza helper bacteria revisited. New
437 Phytologist 176: 22–36.
438 Haas D., Defago G. 2005. Biological control of soil-borne pathogens by fluorescent
439 pseudomonads. Nature Reviews Microbiology 3:307-19.
440 Halsey JA., e Silva MDCP., Andreote FD. 2016. Bacterial selection by mycospheres of
441 Atlantic Rainforest mushrooms. Antonie van Leeuwenhoek 109: 1353–1365.
442 Han MV., Thomas GW., Lugo-Martinez J., Hahn MW. 2013. Estimating gene gain and loss
443 rates in the presence of error in genome assembly and annotation using CAFE 3. Molecular
444 Biology and Evolution 30:1987–1997.

Hibbett DS., Gilbert L-B., Donoghue MJ. 2000. Evolutionary instability of ectomycorrhizal symbioses in basidiomycetes. Nature 407:506–508.
Hibbett D., Grimaldi D., Donoghue M. 1997. Fossil mushrooms from Miocene and Cretaceous ambers and the evolution of Homobasidiomycetes. American Journal of Botany 84:981.
Holt C., Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC bioinformatics 12:491.
Hyatt D., Chen G-L., LoCascio PF., Land ML., Larimer FW., Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.
Janssen PH. 2006. Identifying the Dominant Soil Bacterial Taxa in Libraries of 16S rRNA and 16S rRNA Genes. Applied and Environmental Microbiology 72:1719–1728.
Jones P., Binns D., Chang H-Y., Fraser M., Li W., McAnulla C., McWilliam H., Maslen J., Mitchell A., Nuka G. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240.
Kanehisa M., Sato Y., Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. Journal of Molecular Biology 428:726–731.
Katoh K., Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30:772–780.
Kohler A., Kuo A., Nagy LG., Morin E, Barry KW., Buscot F., Canbäck B., Choi C., Cichocki N., Clum A. 2015. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. Nature Genetics 47:410–415.
Krechel A., Faupel A., Hallmann J., Ulrich A., Berg G. 2002. Potato-associated bacteria and their antagonistic potential towards plant-pathogenic fungi and the plant-parasitic nematode *Meloidogyne incognita* (Kofoid & White) Chitwood. Canadian Journal of Microbiology 48:772–786
Kõljalg U., Nilsson RH., Abarenkov K., Tedersoo L., Taylor AF., Bahram M., Bates ST., Bruns TD., Bengtsson-Palme J., Callaghan TM., Douglas B., *et al*. 2013. Towards a unified paradigm for sequence-based identification of fungi. Molecular ecology. 21:5271–7.
Kumari D., Reddy MS., Upadhyay RC. 2013. Diversity of cultivable bacteria associated with fruiting bodies of wild Himalayan Cantharellus spp. Annals of Microbiology 63: 845-853.
Lackner G., Partida-Martinez LP., Hertweck C. 2009. Endofungal bacteria as producers of mycotoxins. Trends in Microbiology 17:570–576.
Langmead B., Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357–359.
Larsen TS., Krogh A. 2003. EasyGene–a prokaryotic gene finder that ranks ORFs by statistical significance. BMC Bioinformatics 4:21.
Lartillot, N., T. Lepage, and S. Blanquart. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25:2286–2288. Oxford University Press.
Li L., Stoeckert CJ., Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Research 13:2178–2189.
Martin F., Kohler A., Murat C., Balestrini R., Coutinho PM., Jaillon O., Montanini B., Morin E., Noel B., Percudani R. 2010. Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. Nature 464:1033–1038.

490 Matheny PB., Aime MC., Bougher NL., Buyck B., Desjardin DE., Horak E., Kropp BR., Lodge
491 DJ., Soytong K., Trappe JM., Hibbett DS. 2009. Out of the Palaeotropics? Historical
492 biogeography and diversification of the cosmopolitan ectomycorrhizal mushroom family
493 Inocybaceae. Journal of Biogeography 36:577–92.
494 McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K.,
495 Altshuler D., Gabriel S., Daly M. 2010. The Genome Analysis Toolkit: a MapReduce
496 framework for analyzing next-generation DNA sequencing data. Genome Research
497 20:1297–1303.
498 Meyer F., Paarmann D., D'Souza M., Olson R., Glass EM., Kubal M., Paczian T., Rodriguez A.,
499 Stevens R., Wilke A. 2008. The metagenomics RAST server–a public resource for the
500 automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics
501 9:386.
502 Minh BQ., Nguyen MAT., von Haeseler A. 2013. Ultrafast approximation for phylogenetic
503 bootstrap. Molecular biology and evolution 30:1188–1195.
504 Mirarab S., Warnow T. 2015. ASTRAL-II: coalescent-based species tree estimation with many
505 hundreds of taxa and thousands of genes. Bioinformatics 31:i44–i52.
506 Mu LL., Yun YB., Park SJ., Cha JS., Kim YK. 2015. Various pathogenic Pseudomonas strains
507 that cause brown blotch disease in cultivated mushrooms. Journal of Applied Biological
508 Chemistry 58:349–354.
509 Munsch P., Alatossava T., Marttinen N., Meyer JM., Christen R., Garden L. 2002.
510 Pseudomonas costantinii sp. nov., another causal agent of brown blotch disease, isolated
511 from cultivated mushroom sporophores in Finland. International Journal of Systematic and
512 Evolutionary Microbiology 52:1973–1983.
513 Nazir R., Warmink JA., Boersma H., van Elsas JD. 2010a. Mechanisms that promote bacterial
514 fitness in fungal-affected soil microhabitats. FEMS Microbiology Ecology 71, 169–185
515 Nazir R., Boersma FGH., Warmink JA., van Elsas JD. 2010b. *Lyphyllum* sp. strain Karsten
516 alleviates pH pressure in acid soil and enhances the survival of *Variovorax paradoxus* HB44
517 and other bacteria in the mycosphere. Soil Biology and Biochemistry. 42, 2146–2152.
518 Nguyen LT., Schmidt HA., von Haeseler A., Minh B.Q.2014. IQ-TREE: a fast and effective
519 stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular biology
520 and evolution, 32: 268–274.
521 Olson Å., Aerts A., Asiegbu F., Belbahri L., Bouzid O., Broberg A., Canbäck B., Coutinho
522 PM., Cullen D., Dalman K. 2012. Insight into trade-off between wood decay and parasitism
523 from the genome of a fungal forest pathogen. New Phytologist 194:1001–1013.
524 Parra G., Bradnam K., Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in
525 eukaryotic genomes. Bioinformatics 23:1061–1067.
526 Quandt CA., Kohler A., Hesse C., Sharpton T., Martin F., Spatafora JW. 2015. Metagenome
527 sequence of Elaphomycesgranulatus from sporocarp tissue reveals Ascomycota
528 ectomycorrhizal fingerprints of genome expansion and a Proteobacteria rich microbiome.
529 Environmental Microbiology 17:2952–2968.
530 Pent M., Põldmaa K., Bahram M. 2017. Bacterial communities in boreal forest mushrooms are
531 shaped both by soil parameters and host identity. Frontier in Microbiology 8:836.
532 Rangel-Castro JI., Danell E., Taylor AFS. 2002. Use of different nitrogen sources by the edible
533 ectomycorrhizal mushroom Cantharellus cibarius. Mycorrhiza 12:131–137.

534 Rangel-Castro JI., Levenfors JJ., Danell E. 2002. Physiological and genetic characterization of
535 fluorescent Pseudomonas associated with *Cantharellus cibarius*. Canadian Journal of
536 Microbiology 48:739–748.
537 Riedlinger J., Schrey SD., Tarkka MT., Hampp R., Kapur M., Fiedler H-P. 2006. Auxofuran, a
538 novel metabolite that stimulates the growth of fly agaric, is produced by the mycorrhiza
539 helper bacterium Streptomyces strain AcH 505. Applied and Environmental Microbiology
540 72:3550–3557.
541 Rimmer A., Phan H., Mathieson I., Iqbal Z., Twigg SR., Wilkie AO., McVean G., Lunter G.,
542 Consortium W. 2014. Integrating mapping-, assembly-and haplotype-based approaches for
543 calling variants in clinical sequencing applications. Nature Genetics 46:912–918.
544 Smit A., Hubley R. 2011. RepeatModeler-1.0. 5. Institute of Systems Biology.
545 Smith SE., Read DJ. 2010. *Mycorrhizal symbiosis* 3rd edn. Academic Press: London, UK.
546 Spanu PD., Abbott JC., Amselem J., Burgis TA., Soanes DM., Stuber K., Ver Loren van
547 Themaat E, *et al*. 2010. Genome expansion and gene loss in powdery mildew fungi reveal
548 tradeoffs in extreme parasitism. Science 330:1543–1546.
549 Splivallo R., Deveau A., Valdez N., Kirchhoff N., Frey-Klett P., Karlovsky P. 2015. Bacteria
550 associated with truffle-fruiting bodies contribute to truffle aroma. Environmental
551 Microbiology 17:2647–2660.
552 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
553 large phylogenies. Bioinformatics 30:1312–1313.
554 Stopnisek N., Zühlke D., Carlier A., Barberán A., Fierer N., Becher D., Riedel K., Eberl L.,
555 Weisskopf L. 2016. Molecular mechanisms underlying the close association between soil
556 Burkholderia and fungi. The ISME journal 10:253–264.
557 Stubblefield SP., Taylor TN., Beck CB.1985. Studies of paleozoic fungi. IV. Wood-decaying
558 fungi in Callixylon newberryi from the upper Devonian. American Journal of Botany 1765-
559 1774.
560 Talavera G., Castresana J. 2007. Improvement of phylogenies after removing divergent and
561 ambiguously aligned blocks from protein sequence alignments. Systematic Biology 56:564–
562 577.
563 Tedersoo L., May TW., Smith ME. 2010. Ectomycorrhizal lifestyle in fungi: global diversity,
564 distribution, and evolution of phylogenetic lineages. Mycorrhiza 20:217–263.
565 Thorne JL., Kishino H., Painter IS.1998. Estimating the rate of evolution of the rate of
566 molecular evolution. Molecular biology and evolution, 15: 1647–1657.
567 Tian X., Cheng X., Mao Z., Chen G., Yang J., Xie B. 2011. Composition of Bacterial
568 communities associated with plant-parasitic nematode *Bursaphelenchus mucronatus*.
569 Current Microbiology 62:117–125.
570 Torres-Cortés G., Ghignone S., Bonfante P., Schüßler A. 2015. Mosaic genome of endobacteria
571 in arbuscular mycorrhizal fungi: Transkingdom gene transfer in an ancient mycoplasma-
572 fungus association. Proceedings of the National Academy of Sciences 112:7785–7790.
573 Tsukamoto T., Murata H., Shirata A. 2002. Identification of non-pseudomonad bacteria from
574 fruit bodies of wild agaricales fungi that detoxify tolaasin produced by *Pseudomonas*
575 *tolaasii*. Bioscience, Biotechnology, and Biochemistry 66:2201–2208.
576 UniProt Consortium. 2014. UniProt: a hub for protein information. Nucleic acids research.
577 gku989.

Uroz S., Oger P., Morin E., Frey-Klett P. 2012. Distinct ectomycorrhizospheres share similar bacterial communities as revealed by pyrosequencing-based analysis of 16S rRNA genes. Applied and Environmental Microbiology 78:3020–3024.

Veldre V., Abarenkov K., Bahram M., Martos F., Selosse M-A., Tamm H., Kõljalg U., Tedersoo L. 2013. Evolution of nutritional modes of Ceratobasidiaceae (Cantharellales, Basidiomycota) as revealed from publicly available ITS sequences. Fungal Ecology 6:256–268.

Vicente CSL., Nascimento F., Espada M., Barbosa P., Mota M., Glick BR., Oliveira S. 2012. Characterization of bacteria associated with pinewood nematode *Bursaphelenchus xylophilus*. PLOS ONE 7: e46661. https://doi.org/10.1371/journal.pone.0046661.

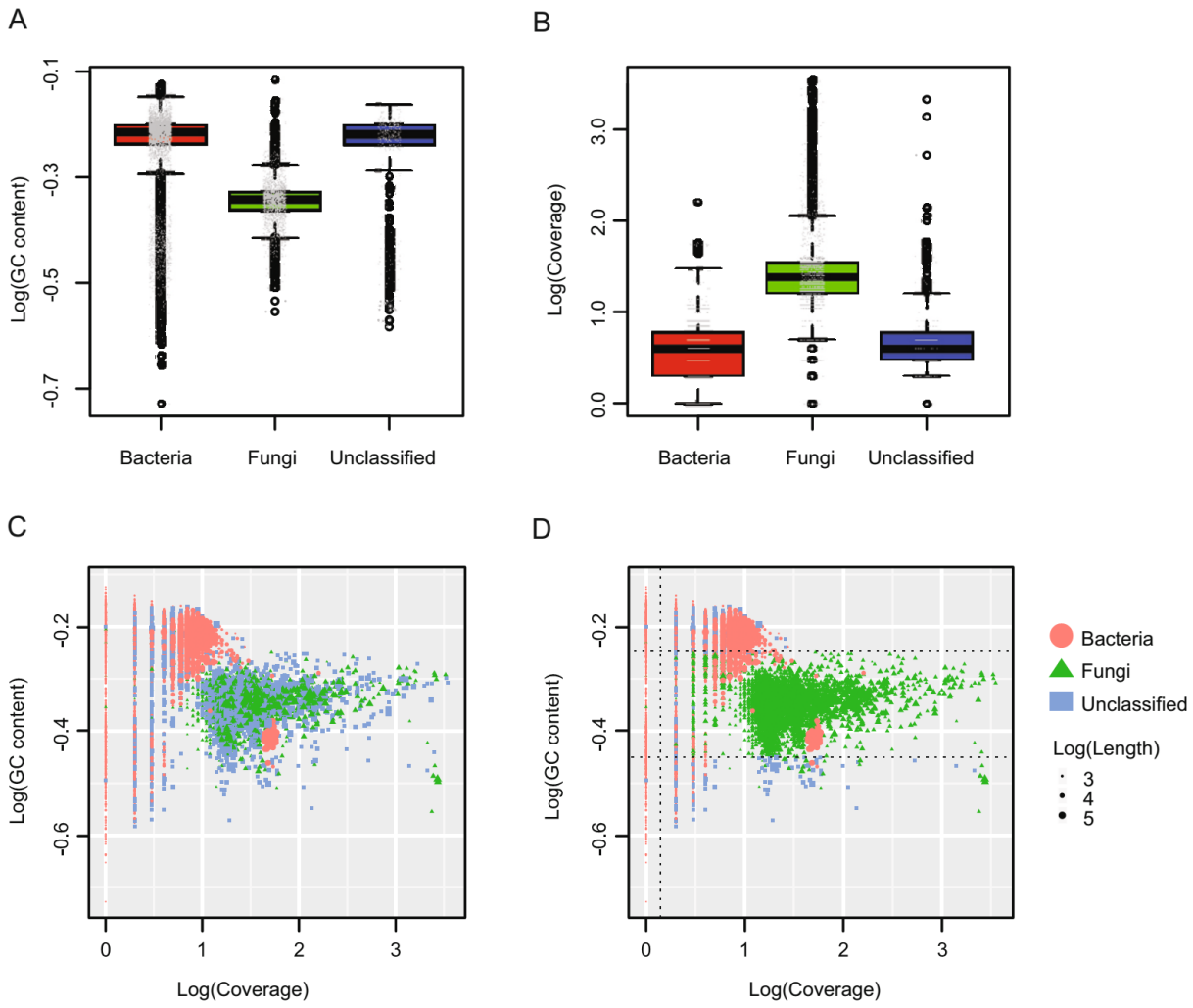
Wang HC., Minh BQ, Susko E., Roger AJ. 2017. Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. Systematic Biology, doi: 10.1093/sysbio/syx068.

Warmink JA., Nazir R., van Elsas JD. 2009. Universal and species-specific bacterial "fungiphiles" in the mycospheres of different basidiomycetous fungi. Environmental Microbiology 11:300–312.

Yin Y., Mao X., Yang J., Chen X., Mao F., Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Research 29:40:W445–51.

612 Figures

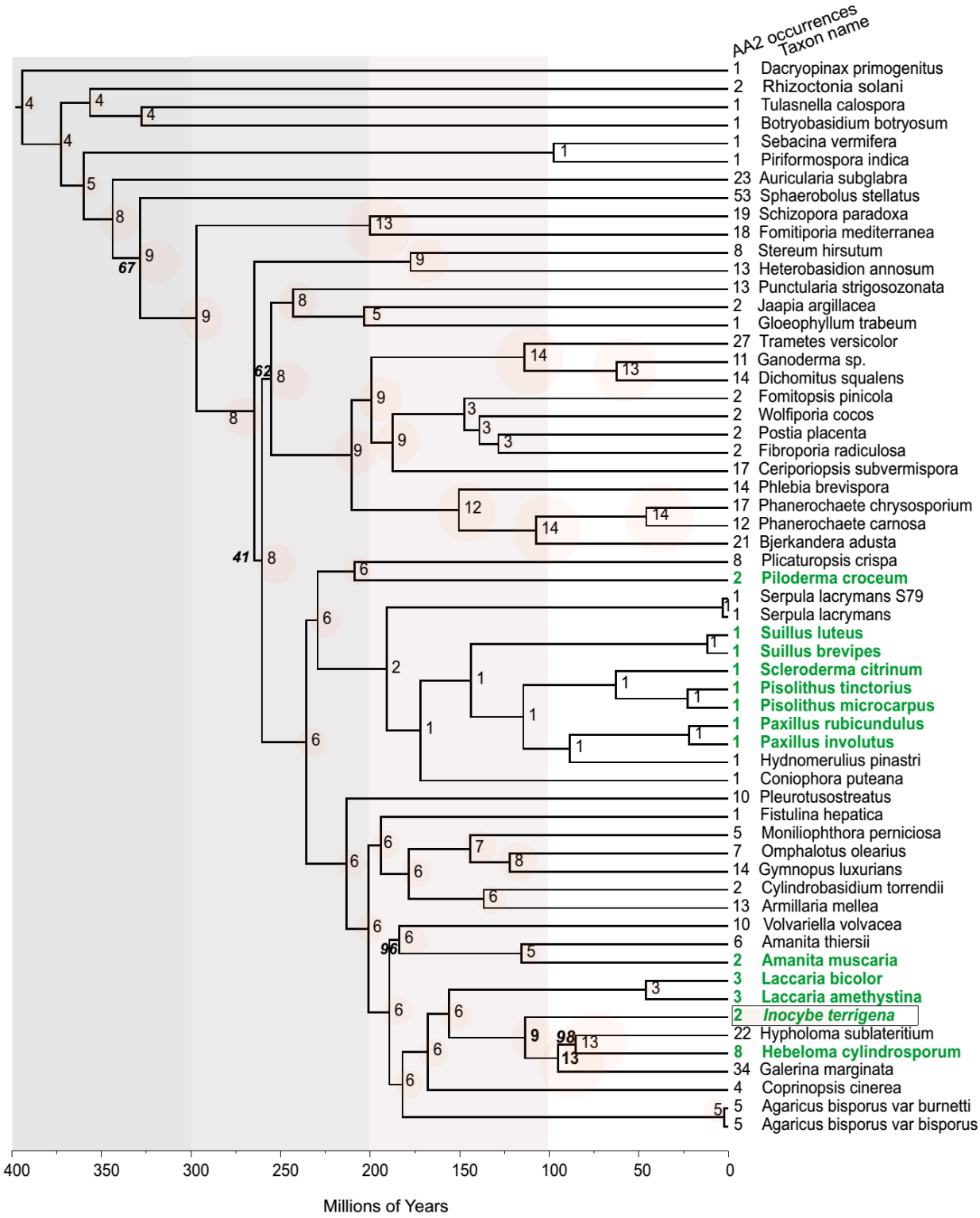
613



614

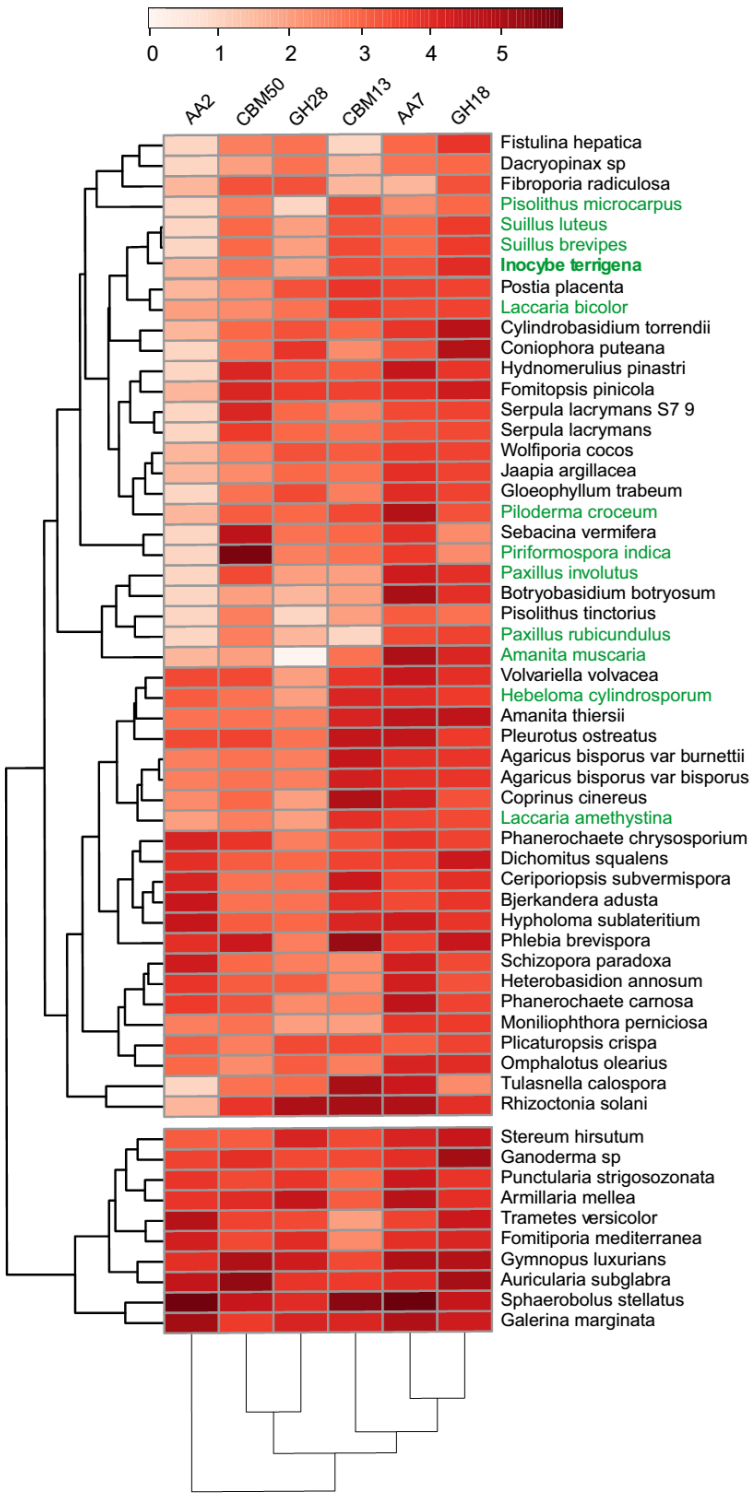
615 **Figure 1**. Bacterial and fungal assemblies can be separated based on GC content and coverage.
616 (A-B) The boxplots of GC content (A) and median coverage (B) for bacterial, fungal contigs
617 identified based on Blast searches. (C) The scatterplot of GC content as a function of median
618 coverage of contigs. (D) Same as D however unclassified contigs with median coverage >1 and
619 GC content >0.35 and <0.56 were classified as fungi. In C and D, unclassified contigs of <1000
620 bp length have been excluded.

621


AA2 occurrences
Taxon name
1 Dacryopinax primogenitus
2 Rhizoctonia solani
1 Tulasnella calospora
1 Botryobasidium botryosum
1 Sebacina vermifera
1 Piriformospora indica
23 Auricularia subglabra
53 Sphaerobolus stellatus
19 Schizopora paradoxa
18 Fomitiporia mediterranea
8 Stereum hirsutum
13 Heterobasidion annosum
13 Punctularia strigosozonata
2 Jaapia argillacea
1 Gloeophyllum trabeum
27 Trametes versicolor
11 Ganoderma sp.
14 Dichomitus squalens
2 Fomitopsis pinicola
2 Wolfiporia cocos
2 Postia placenta
2 Fibroporia radiculosa
17 Ceriporiopsis subvermispora
14 Phlebia brevispora
17 Phanerochaete chrysosporium
12 Phanerochaete carnosa
21 Bjerkandera adusta
8 Plicaturopsis crispa
2 Piloderma croceum
1 Serpula lacrymans S79
1 Serpula lacrymans
1 Suillus luteus
1 Suillus brevipes
1 Scleroderma citrinum
1 Pisolithus tinctorius
1 Pisolithus microcarpus
1 Paxillus rubicundulus
1 Paxillus involutus
1 Hydnomerulius pinastri
1 Coniophora puteana
10 Pleurotusostreatus
1 Fistulina hepatica
5 Moniliophthora perniciosa
7 Omphalotus olearius
14 Gymnopus luxurians
2 Cylindrobasidium torrendii
13 Armillaria mellea
10 Volvariella volvacea
6 Amanita thiersii
2 Amanita muscaria
3 Laccaria bicolor
3 Laccaria amethystina
2 Inocybe terrigena
22 Hypholoma sublateritium
8 Hebeloma cylindrosporum
34 Galerina marginata
4 Coprinopsis cinerea
5 Agaricus bisporus var burnetti
5 Agaricus bisporus var bisporus
400 350 300 250 200 150 100 50 0
Millions of Years

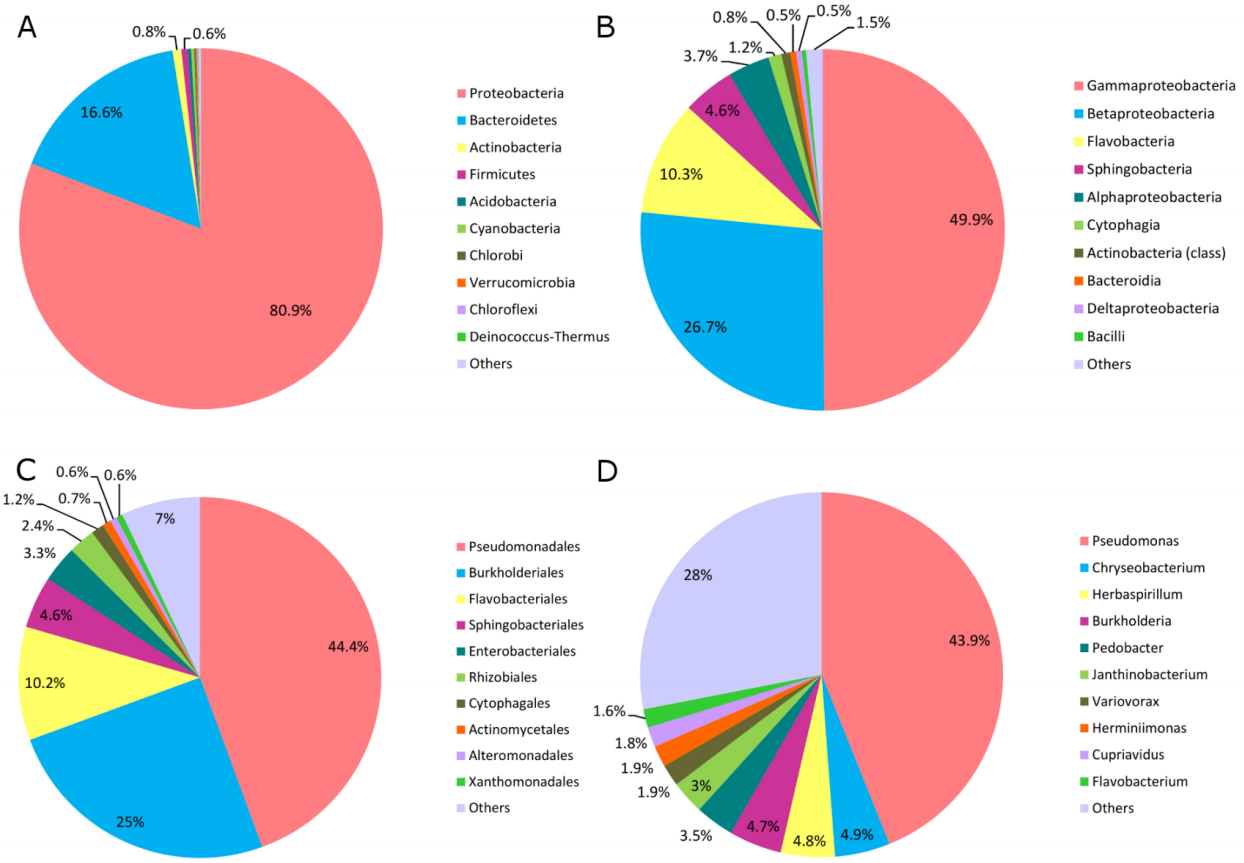
622

623 **Figure 2**. Phylogenetic tree showing the phylogenetic relationship of *Inocybe terrigena* and
624 other published ectomycorrhizal and saprotrophic fungi. All clades have a support value of 100%
625 except those that are indicted in italic text. Non-italic node labels and numbers next to taxon
626 names represent the number of AA2 gene occurrences for each parent node and taxon, indicating
627 gain/loss of AA2 genes in each taxon. Note that *I. terrigena* belongs to a clade which has gained
628 three AA2 genes from its parent node, but it has lost seven AA2 genes compared to other
629 members of this clade.


0
1
2
3
4
5
AA2
CBM50
GH28
CBM13
AA7
GH18
Fistulina hepatica
Dacryopinax sp
Fibroporia radiculosa
Pisolithus microcarpus
Suillus luteus
Suillus brevipes
Inocybe terrigena
Postia placenta
Laccaria bicolor
Cylindrobasidium torrendii
Coniophora puteana
Hydnomerulius pinastri
Fomitopsis pinicola
Serpula lacrymans S7 9
Serpula lacrymans
Wolfiporia cocos
Jaapia argillacea
Gloeophyllum trabeum
Piloderma croceum
Sebacina vermifera
Piriformospora indica
Paxillus involutus
Botryobasidium botryosum
Pisolithus tinctorius
Paxillus rubicundulus
Amanita muscaria
Volvariella volvacea
Hebeloma cylindrosporum
Amanita thiersii
Pleurotus ostreatus
Agaricus bisporus var burnettii
Agaricus bisporus var bisporus
Coprinus cinereus
Laccaria amethystina
Phanerochaete chrysosporium
Dichomitus squalens
Ceriporiopsis subvermispora
Bjerkandera adusta
Hypholoma sublateritium
Phlebia brevispora
Schizopora paradoxa
Heterobasidion annosum
Phanerochaete carnosa
Moniliophthora perniciosa
Plicaturopsis crispa
Omphalotus olearius
Tulasnella calospora
Rhizoctonia solani
Stereum hirsutum
Ganoderma sp
Punctularia strigosozonata
Armillaria mellea
Trametes versicolor
Fomitiporia mediterranea
Gymnopus luxurians
Auricularia subglabra
Sphaerobolus stellatus
Galerina marginata

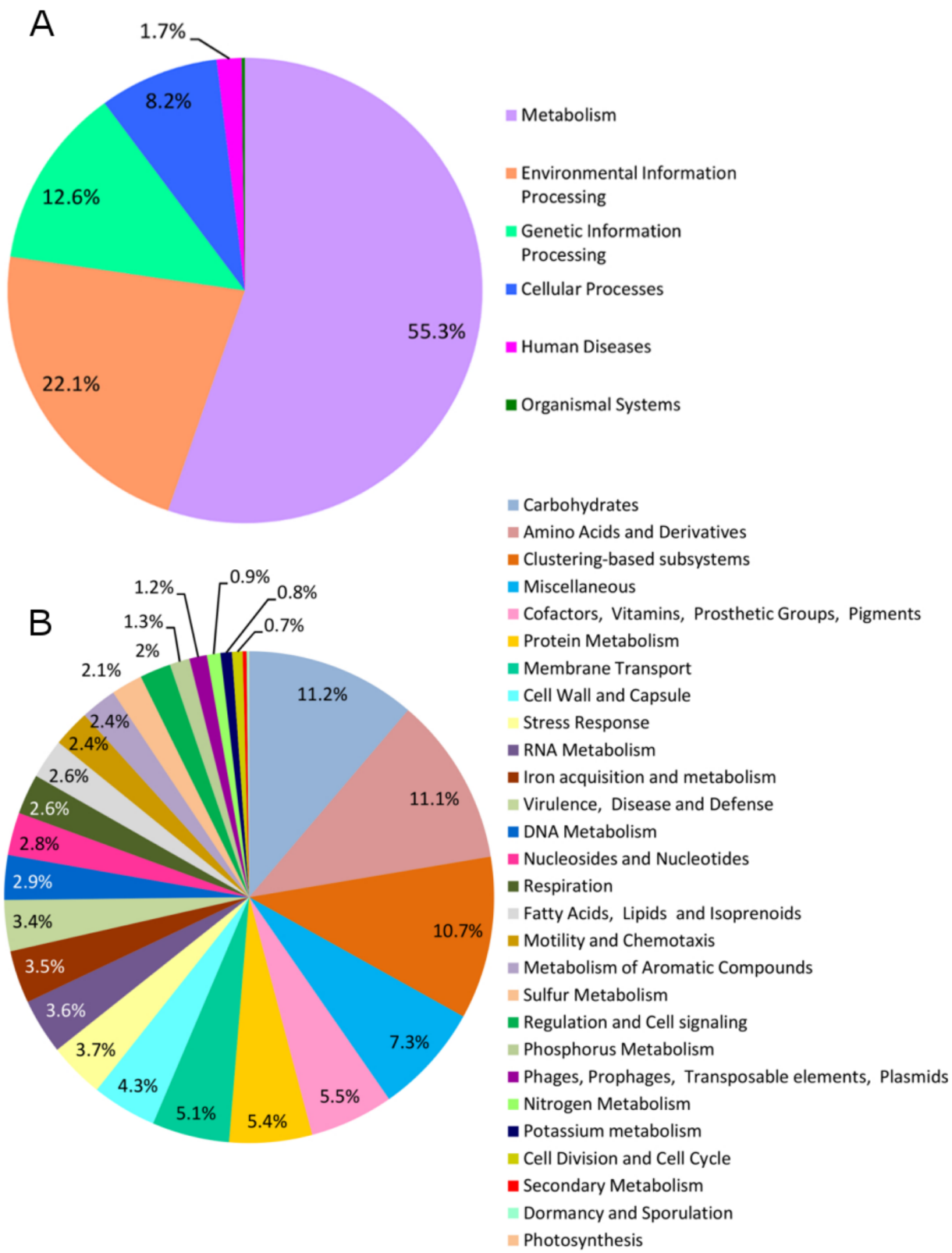
630

631 **Figure 3**. Heatmap of six CAZyme families that showed significant expansion or contraction
632 across 59 analysed genomes in this study.



633

634 **Figure 4**. Pie chart showing the relative abundance of 10 most common bacterial taxa at phylum
635 (A), class (B), order (C) and genus (D) level in *Inocybe terrigena* fruitbody based on
636 representative hits of RefSeq database (at e-value<1×10^{-5}, %identity>60) using MG-RAST.
637 Bacterial groups with abundance ≥ 0.5% are presented. All fungal, bacterial and unclassified
638 contigs were included.


A
1.7%
8.2%
12.6%
55.3%
22.1%
Metabolism
Environmental Information Processing
Genetic Information Processing
Cellular Processes
Human Diseases
Organismal Systems
B
1.2%
0.9%
0.8%
1.3%
0.7%
2%
2.1%
11.2%
2.4%
2.4%
2.6%
11.1%
2.6%
2.8%
2.9%
3.4%
10.7%
3.5%
3.6%
7.3%
3.7%
4.3%
5.5%
5.1%
5.4%
Carbohydrates
Amino Acids and Derivatives
Clustering-based subsystems
Miscellaneous
Cofactors, Vitamins, Prosthetic Groups, Pigments
Protein Metabolism
Membrane Transport
Cell Wall and Capsule
Stress Response
RNA Metabolism
Iron acquisition and metabolism
Virulence, Disease and Defense
DNA Metabolism
Nucleosides and Nucleotides
Respiration
Fatty Acids, Lipids and Isoprenoids
Motility and Chemotaxis
Metabolism of Aromatic Compounds
Sulfur Metabolism
Regulation and Cell signaling
Phosphorus Metabolism
Phages, Prophages, Transposable elements, Plasmids
Nitrogen Metabolism
Potassium metabolism
Cell Division and Cell Cycle
Secondary Metabolism
Dormancy and Sporulation
Photosynthesis

639

640 **Figure 5.** Pie chart showing the relative abundance of bacterial gene functional categories in
641 *Inocybe terrigena* fruitbody based on KEGG Orthology groups **(A)** and Subsystems **(B)**.
642 Functional categories with relative abundance of ≥ 0.5% are presented. Only contigs that were
643 initially identified as bacteria based on BLAST searches were included.

644