PeerJ Preprints NOT PEER-REVIEWED

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 ectomycorrhizal 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 supporting 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*

29

30 **Introduction**

31 Ectomycorrhizal fungi constitute a major component of fungal communities in terrestrial
32 ecosystems, functioning to facilitate nutrient uptake and carbon storage by plants (Smith and
33 Read 2010). Evidence from phylogenetic studies, particularly ribosomal RNA genes, suggests
34 that ectomycorrhizal fungi have evolved independently in several fungal lineages (Hibbett *et al.,*
35 2000; Tedersoo *et al.,* 2010; Veldre *et al.,* 2013) but the genomic adaptations associated with
36 these functional shifts are only beginning to be characterized (Kohler *et al.,* 2015).

37 Recent genomic studies focusing on fungi and using high-throughput sequencing (HTS)
38 technologies have improved our understanding of the evolution of mycorrhizal symbiosis. These
39 studies provide evidence that while mycorrhizal fungi have retained much of their enzymatic
40 capacity to release nutrients from complex organic compounds, certain carbohydrate active
41 enzyme (CAZyme) families have contracted during the evolution of mycorrhizal fungi compared
42 to their saprotrophic ancestors (Martin *et al.,* 2010; Kohler *et al.,* 2015). In addition, metabolite
43 pathways and secondary metabolites vary between obligate biotrophs and saprotrophs. This
44 adaptation reflects a shift in life history, contributing to variation in genome size, characteristic
45 reductions in gene families such as transporters and plant cell wall degradation enzymes (Martin
46 *et al.,* 2010; Spanu *et al.,* 2010; Floudas *et al.,* 2012; Kohler *et al.,* 2015). Ectomycorrhizal fungi
47 possess a repertoire of genes encoding cellulose degrading enzymes to help release simple
48 organic compounds that are available for uptake by plants, which substantially accelerates soil
49 nutrient cycling (Martin *et al.,* 2010). So far, most fungal genome sequencing studies have
50 employed culture-dependent techniques, which omit major ectomycorrhizal lineages that are
51 difficult to culture.

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

58 2013) and mycorrhization (Frey-Klett *et al.,* 2007; Aspray *et al.,* 2013), yet we know little about
59 the associated bacterial taxa and functions in epigeous fruiting-bodies.

60 We performed metagenomic sequencing on DNA extracted from fruiting-body tissues to
61 characterize the genomes of the dikaryotic fungus *Inocybe terrigena* (Inocybaceae, Agaricales)
62 and those of its associated bacterial community. Inocybaceae is a diverse ectomycorrhizal fungal
63 lineage that is thought to have evolved independently from other ectomycorrhizal lineages
64 (Matheny *et al.,* 2009). Nevertheless, there is no published work on the evolution of mycorrhizal
65 status in Inocybaceae using comparative genomics. We compared all CAZyme gene families for
66 the genomes of *I. terrigena* and and 58 other genomes from the Agaricomycetidae to look for
67 significant expansion or contraction in gene families across multiple, independent evolutionary
68 transitions to mycorrhizal symbiosis (Matheny *et al.,* 2009). Additionally, we characterized the
69 taxonomic composition and potential functions of the associated bacteria in fruiting-body tissues.
70

71 **Material and Methods**

72 DNA was extracted from a dried collection of *Inocybe terrigena* (field accession no: MR00339;
73 coordinates: N59°0'7.52, E17°42'2.42"; date: 2013-09-20; herbarium: UPS), collected and
74 identified by MR, using Plant mix DNeasy DNA Isolation kit. To obtain DNA of sufficient
75 quantity, four DNA extractions from lamellae of the fruiting-body were pooled. Lamellae are
76 internal spore bearing layered structures in fruiting-bodies and are therefore expected to provide
77 higher DNA per unit of material. In addition, as lamellae are protected during development they
78 have a lower risk of contamination than other exposed parts of the fruiting-body. We took extra
79 care to minimize contamination during sample collecting, storing and handling. PCR free,
80 paired-end, 300bp insert libraries were constructed from total DNA and sequenced on 1/10 of an
81 Illumina HiSeq 2500 lane at Sci-Life laboratory (Stockholm, Sweden). Raw Illumina reads,
82 available at the Sequence Read Archive (SRA) under accession number SRP066410, were
83 quality filtered and trimmed using Fastx Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/[2]) using
84 the following settings: -q 25 -p 90 and -t 25 -l 20, respectively. Genome assembly was performed
85 using Spades de novo (Bankevich *et al.,* 2012) with default settings. To sort assembled contigs
86 into fungi and bacteria, we used BLAST queries to compare contigs to previously published
87 whole genomes of bacteria and fungi downloaded from GenBank and the Joint Genome Institute

88 (JGI; www.jgi.doe.gov) website, respectively. To evaluate the accuracy and coverage of the
89 resulting assemblies, we mapped post-filtered reads to the assembled contigs using Bowtie 2
90 (Langmead *et al.,* 2012). We used median coverage in combination with GC content to
91 determine whether contigs not classified by BLAST were of bacterial or fungal origin, by
92 comparison to the values of contigs classified by BLAST. Unclassified contigs shorter than 1000
93 bp were excluded from the analysis, as GC content and median coverage varied greatly for these
94 and taxonomic assignment was deemed unreliable. Bacterial and eukaryotic species present in
95 the final assemblies were identified using (previously constructed) HMMer profiles to recognize
96 and trim ribosomal operons and ITS region, respectively (Lagesen *et al.,* 2007; Bengtsson-Palme
97 *et al.,* 2013). Additionally, we used the MG-RAST pipeline (Meyer *et al.,* 2008) to infer both
98 taxonomic and functional annotations based on assemblies.

99 Protein annotation of fungal and bacterial assemblies was performed using the
100 Maker2.3.36 pipeline (Holt and Yandell 2011) and Prodigal (Hyatt *et al.,* 2010) with default
101 parameters, respectively. Core Eukaryotic Mapping Genes Approach (CEGMA) (Parra *et al.,*
102 2008) was used to evaluate genome completeness and to generate preliminary annotations as
103 training sets in Maker2. RepeatModeler (Smit and Hubley 2011) was used to generate a
104 classified repeat library for the metagenome assembly. This repetitive sequence library was
105 combined with a Maker2 provided a transposable element library and was used in RepeatMasker
106 3.0 (Smit *et al.,* 2010) for masking within the Maker2 pipeline. Additionally, the proteomes of
107 *Laccaria bicolor*, *Coprinopsis cinerea* and the UniProt reference proteomes (The UniProt
108 Consortium, 2014) were used as protein evidence to generate hints for *ab-initio* predictors. Three
109 rounds of training were used for the *ab-initio* gene predictors SNAP and Augustus while self-
110 training was used for GeneMark-ES. Overcalling genes is common for *ab-initio* gene predictors
111 (Larsen and Krogh 2003) so "keep_preds=0" was set in Maker2 to only call gene models which
112 were supported by protein evidence (AED<1). Predicted proteins lacking functional evidence
113 were scanned for protein family domains (PFAM) using InterProScan5 (IPRscan5) (Jones *et al.,*
114 2014), those containing PFAM domains were promoted to full gene models using scripts
115 included with Maker2. Protein models were assigned to multigene families using the OrthoMCL
116 software (v2.0.9) (Li *et al.,* 2003). BLASTX was used to compare and sort bacterial and fungal
117 protein models against the whole NCBI protein database. Analyses were facilitated with a
118 custom perl pipeline, which is available in the supplementary information. Fungal protein models

119 were compared to non-redundant genomes using BlastKOALA (Kanehisa *et al.,* 2016) and
120 eggNOG-mapper (Huerta-Cepas *et al*., 2017) for functional annotation with additional screening
121 for carbohydrate-active enzymes (CAZYmes) using the dbCAN pipeline (Yin *et al.,* 2012).

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

127 Phylogenetic analysis was performed based on complementary supermatrix and supertree
128 approaches with 74 publicly available Agaricomycetes genomes (retrieved from the JGI
129 database). Single-copy genes were identified based on MCL clusters on BLAST e-values with
130 the inflation parameter set to 2.0, in an additional quality check they were clustered using
131 OrthoMCL with only those single copy genes present in 75% of taxa used for phylogenetic
132 analysis. Amino acid sequences were aligned using MAFFT 7 (Katoh and Standley 2013) with
133 highly variable regions removed using Gblocks (Talavera *et al.,* 2007) with the settings t=d and
134 b5=h. The resulting concatenated alignment was used to estimate a maximum likelihood
135 phylogeny with 1,000 UltraFast (UF) bootstrap replicates (Minh *et al.,* 2013) using IQ-Tree 1.5.5
136 (Nguyen *et al.,* 2015). Sequence evolution was modeled using the Posterior Mean Site Frequency
137 (PSMF) model (Wang *et al.,* 2017) and 60 mixture classes. PhyloBayes 3.3 (Lartillot *et al.,*
138 2009) with site-specific evolutionary rates modeled using non-parametric infinite mixtures
139 (CAT-GTR) was used to generate a fossil calibrated ultrametric tree from our whole genome
140 concatenated tree. A diffuse gamma prior with the mean equal to the standard deviation was used
141 for the root of the tree. A log-normal auto-correlated relaxed clock model (Thorne *et al.,* 1998)
142 with a Dirichlet prior on divergence times was used for dating the tree. Two fossils were
143 employed to calibrate node times to geological time. The first corresponds to a minimum age of
144 90My for the Agaricales (Hibbett *et al.,* 1997). The second corresponds to a minimum age of
145 360My (Stubblefield *et al.,* 1985) for the Basidiomycota. Trees for individual genes were
146 constructed using the PROTGAMMAWAG model and 100 bootstraps in RAxML 8.2.4
147 (Stamatakis 2014). The gene trees were used to estimate to species trees using ASTRAL; support
148 values were calculated using the bootstrapped trees (Mirarab and Warnow 2015). Significant

149 expansion or contraction of CAZyme gene families was determined for a reduced phylogeny of
150 the 59 taxa belonging to Agaricomycotina using CAFE 3.1 (Han *et al.,* 2013).

151

152 **Results**

153 *Genomic features of Inocybe terrigena*

154 After quality filtering and trimming, 29,244,774 paired end reads (96%) were used for assembly.
155 Statistics for sequencing and assembly are available in Table S1. BLAST searches were
156 performed with the resulting 31,471 assembled contigs against reference genomes resulted in
157 2,262 contigs (N50: 30,623, total length: 26,127,495 bp) and 11,596 contigs (N50: 8,777, total
158 length 34,853,629 bp) initially identified as fungal and bacterial, respectively. Mitochondrial
159 contigs were determined based on coverage (>1200X) and GC content (<0.40), resulting in 5
160 contigs, with a combined length of 67 kb. In total, 85.8% of non mitochondrial fungal sequences
161 had a GC content between 0.4 and 0.5 (Table S2). Most bacterial contigs are outside this interval
162 (97.2%). Among non-fungal genera, *Pedobacter*, *Pseudomonas* and *Burkholderia* had the
163 highest coverages of 40.24±20.64, 5.73±3.24 and 4.23±1.99, respectively (Fig. S2). In total,
164 56.5% and 10.6% of total bacterial assembly length belonged to *Pseudomonas* and *Pedobacter*,
165 with GC content of 0.608±0.032 and 0.396±0.039, respectively (Fig. 1; Fig. S2). The remaining
166 contigs of bacterial origin had GC content of 0.500±0.115 and coverage of 18.86±108.85 (Fig. 1;
167 Fig. S2). Based on this we classified previously unclassified contigs with median coverage of >1
168 and GC content >0.35 and <0.56 (Fig. 1) as fungal, but made no further attempts classifying
169 contigs as bacteria. The final assembly included 5409 fungal contigs (N50: 19,600, total length:
170 37,436,744 bp) and 12,040 bacterial contigs (N50, 9,606, total length 39,035,878 bp; Table S2).

171 The MGRAST analysis revealed that 77.42% of contigs were of bacterial origin followed
172 by 22.47% fungal, 0.04% viral and 0.03% archaeal contigs. Mapping all reads to the assembly
173 resulted in a total of 28,632,347 mapped reads (93% of all reads), with a median coverage of
174 35X in the fungal contigs (22,274,636 reads; 78% of mapped reads) and 5X for bacterial contigs
175 (5,072,696 reads; 18%). Variant calling revealed 3.4 variant sites per kb in the fungal contigs.

176 BLAST analyses of extracted ITS sequences against UNITE (Kõljalg *et al.,* 2013)
177 resulted in one ITS sequence matching *Inocybe terrigena* (accession number: JF908091;

178 identity=99%, e-value=1e-137, median coverage=2335, contig length=8222), and five additional
179 ITS sequences matching other Eukaryotes. All of the additional eukaryotic contigs including ITS
180 had very low coverage: *Hypomyces odoratus* (identity=100%, e-value=1e-86, median
181 coverage=2, contig length=745), *Selaginella deflexa* (identity=98%, e-value=6e-102, median
182 coverage=2, contig length=1175), *Drosophila subauraria* (identity=84%, e-value=3e-23, median
183 coverage=1, contig length=438), *Cucumis melo* (identity=100%, e-value=3e-04, median
184 coverage=3, contig length=1041). These results indicate a negligible contribution of other
185 Eukaryotes to the resulting *I. terrigena* assembly.

186 The Maker2 gene annotation of fungal (n=2261), unclassified (contigs matching neither
187 fungi nor bacteria; n=17196) and ambivalent (contigs matching both fungi and bacteria; n=417)
188 contigs resulted in 11,918 fungal genes identified (coverage=50.58±142.49, median±SD; GC
189 content=0.467±0.0295). The Prodigal annotation of bacterial contigs resulted in 63,328 bacterial
190 genes. Of these, 3,289 additional fungal genes were identified based on BLAST searches
191 (median coverage=57.33±152.85; GC content=0.471±0.043), indicating some fungal contigs
192 were misidentified as bacterial contigs. Thus, we confirmed the identification of all classified
193 contigs based on BLAST searches using annotated genes against the NCBI protein database, as
194 implemented in the pipeline (Table S2; Supporting Information 1). Based on genes annotated by
195 Prodigal and genes identified from unclassified contigs using Maker2 (n=8190, median
196 coverage= 10.70±11.77; GC content=0.599±0.065), 62,023 genes were identified as bacterial
197 (11,188 and 50,835 by Maker2 and Prodigal, respectively; median coverage=10.07±13.24; GC
198 content=0.588±0.080; Table S3). After this additional round of filtering, 2,855 genes remain
199 unclassified (median coverage=6.36±37.47; GC content=0.598±0.073).

200 In total, the *I. terrigena* genome included 15,207 genes, which is lower than those
201 reported for most ectomycorrhizal fungi (Fig. S1; Table S5). Of 15,207 genes, 24.6% genes were
202 with Pfam domains based on InterPro domain assignments. Using KEGGmapper, 4,467 out of
203 15,207 fungal genes were functionally annotated (29.4%; Fig. S3; Table S3). Using CAZY
204 pipeline, we identified 396 Pfam domains from CAZy Families, including 46 auxiliary activities
205 (AA, 8 families), 33 carbohydrate-binding modules (CBM, 11 families), 47 carbohydrate
206 esterases (CE, 7 families), 168 glycoside hydrolases (GH, 31 families), 94 glycosyltransferases
207 (GT, 34 families) and 8 polysaccharide lyases (PL, 4 families).

208 The CEGMA analysis indicated 91% of 242 full-length, core eukaryotic genes were
209 present in our fungal assembly and 93% were present when partial and full-length alignments
210 were considered. (Table S4).

211 *Comparative genomics*

212 OrthoMCL clustering of 59 Agaricomycetes whole proteomes resulted in 870 single-copy genes
213 present in >75% of taxa, and a concatenated alignment of 148,316 amino acid positions. The
214 resulting Maximum Likelihood phylogenetic tree (Fig. 2) is largely consistent with previous
215 phylogenies (e.g. Kohler *et al.,* 2015) and the ASTRAL phylogeny (Supplementary Information
216 2) had well supported parts. Our CAFE analysis of CAZYmes indicated that 6 of ~220 CAZY
217 families found were significantly expanded or contracted in the 59 genomes analysed here (Fig.
218 3; Table S6). Cluster analysis revealed two major groups of fungi based on CAZyme profiles,
219 and it placed *I. terrigena* in the same cluster as other ectomycorrhizal taxa (Fig. 2). Compared to
220 other genomes included in our comparative genomics analysis, *I. terrigena* has no significantly
221 expanded CAZyme families; however, it is significantly reduced in an important lignin
222 degrading CAZyme family (AA2), which contains class II lignin-modifying peroxidases. While
223 *I. terrigena* belongs to a clade which has gained 3 AA2 from its parent node, it has lost 7 AA2
224 compared to other members of this clade.

225

226 *Microbiome of Inocybe terrigena*

227 Proteobacteria and Bacteroidetes comprised 80.9% and 16.6% of the taxonomic composition in
228 bacterial metagenomics data (Fig. 4A). Gammaproteobacteria (49.9%), Betaproteobacteria
229 (26.7%), Flavobacteria (10.3%), Sphingobacteria (4.6%) and Alphaproteobacteria (3.7%) were
230 the dominant bacterial classes (Fig. 4B). *Pseudomonas*, *Chryseobacterium, Herbaspirillum*,
231 *Burkholderia* and *Pedobacter* were the dominant genera (Fig. 4D). In total, the α-diversity of our
232 metagenome was 63 species based on MG-RAST pipeline, with three dominant species
233 *Pseudomonas fluorescens*, *P. syringae* and *Herbaspirillum seropedicae* contributing 61% to the
234 total bacterial genes. The bacterial contigs contained 28,288 predicted proteins that mostly
235 matched genes with known functions including *metabolism* (55.3%), *information storage and*
236 *processing* (22.1%) and *cellular processes and signaling* (8.2%) (Fig. S4).

237

238 **Discussion**

239 *Genomic features of Inocybe terrigena*

240 Our study is the first to simultaneously analyse the genome and bacterial communities of a
241 dikaryotic fungus. Using one tenth of an Illumina lane, we obtained a genome with a comparable
242 completeness with those reported in previous studies (Kohler *et al.,* 2015; Quandt *et al.,* 2015),
243 indicating the acceptable performance of this approach to investigate the genomes of
244 unculturable fungi based on epigeous fruiting-body tissues. Such analysis is facilitated by
245 significant differences in length, GC content and median coverage between bacterial and fungal
246 genomes, which enable classification of assemblies into fungal or bacterial origin. The GC
247 content of the *I. terrigena* genome (0.47) is in the same range as other ectomycorrhizal species in
248 the Agaricales such as *Laccaria bicolor* (0.47) and *L. amethystina* (0.47), *Hypholoma*
249 *sublateritum* (0.51), *Galerina marginata* (0.48), and *Hebeloma cylindrosporum* (0.48). The GC
250 content of all major bacterial groups except three Bacteroidetes (comprising 5.4% of contigs)
251 were outside this range (Table S2). Together, the contrasting GC content and coverage can be
252 used to accurately remove possible contamination (Laetsch *et al*., 2017) and separate bacterial
253 and fungal contigs, which can reduce the computation cost and database biases, particularly
254 when specific genes are targeted.

255 Our analysis revealed that *I. terrigena* has a smaller proteome based on the number of
256 gene models compared to the previously sequenced ectomycorrhizal Basidiomycetes, e.g.
257 *Laccaria bicolor and Hebeloma cylindrosporum* (Fig. S1, Table S5, Martin *et al.,* 2010; Kohler
258 *et al.,* 2015). Our comparative analysis across a set of available fungal genomes revealed a
259 similar CAZyme profile of *I. terrigena* compared to brown rot and ectomycorrhizal lineages
260 (Fig. 2). Given the significant reduction in certain CAZyme families as a common pattern in
261 genome evolution of ectomycorrhizal fungi (Kohler *et al.,* 2015), the increase in the AA2 family
262 in the lineage leading to the most recent common ancestor to *Inocybe* and its sister clade
263 indicates that this line was not ectomycorrhizal; the significant reduction in AA2 in the lineage
264 leading to *Inocybe* therefore supports a separate origin of the ectomycorrhizal nutritional mode in
265 this lineage (Matheny et al. 2006).

266 *Structure and function of associated bacteria of Inocybe terrigena*

267 Proteobacteria and Bacteroidetes were the most abundant phyla in *I. terrigena,* constituting the
268 largest fraction of bacterial assembly (Fig. S3). Similarly to our study, Proteobacteria has been
269 identified as a very common bacterial group in several ascomycetous (Quandt *et al.,* 2015;
270 Benucci and Bonito 2016; Barbieri *et al.,* 2005, 2007) as well as basidiomycetous (Kumari *et al.,*
271 2013; Pent *et al.,* 2017) fruiting-bodies. There is also evidence that the relative abundance of
272 Proteobacteria is higher in the mycosphere (Warmink *et al.,* 2009) and ectomycorhizosphere
273 (Uroz *et al.,* 2012; Antony-Babu *et al.,* 2013) compared to the surrounding soil, suggesting the
274 tendency of Proteobacteria for colonizing fungus-related habitats. A high relative abundance of
275 Bacteroidetes, the second-largest bacterial phylum in *I. terrigena*, is also often found in
276 ectomycorhizosphere, mycosphere and fruiting-bodies of ascomycetous as well as
277 basidiomycetous fungi (Uroz *et al.,* 2012; Antony-Babu *et al.,* 2013; Benucci and Bonito 2016;
278 Halsey *et al.,* 2016; Pent *et al.,* 2017). In particular, the relative abundance of Sphingobacteria in
279 *I. terrigena* was similar to that in *Elaphomyces granulatus* (Quandt *et al.,* 2015). Similarly to *I.*
280 *terrigena* fruiting-body Acidobacteria, Actinobacteria, Verrucomicrobia, Firmicutes and
281 Cyanobacteria form a small fraction of the bacterial community in basidiomycetous (Pent *et al.,*
282 2017) as in ascomycetous (Antony-Babu *et al.,* 2013) fruiting-bodies, whereas they are highly
283 represented in soil (Eilers *et al.,* 2010; Bergmann *et al.,* 2011). The observed level of specificity
284 of fungal associated bacteria may be related to close fungal-bacterial interactions or fungal-
285 driven changes in habitat conditions, such as change in pH or nutrient conditions in closely
286 related bulk soil (Danell *et al.,* 1993; Nazir *et al.,* 2010a,b).

287 The most abundant bacterial genera associated with *I. terrigena* are known to have
288 symbiotic functions in fungal tissues. Although major bacterial taxa found in *I. terrigena* (Fig.
289 S3) are common in soil (Baldani *et al.,* 1986; Janssen 2006), the genus *Pseudomonas* is also one
290 of the most frequently identified bacterial groups in basidiomycetous fruiting-bodies and
291 ectomycorrhizas (Deveau *et al.,* 2007; Kumari *et al.,* 2013). While *Pseudomonas* strains are
292 mainly saprotrophic, they are also known to alleviate detrimental effect of pathogens on plant
293 roots and leaves (Haas and Defago 2005) and facilitate mycorrhizal establishment *(*Dominguez *et*
294 *al.,* 2012*).* Some strains are also plant pathogenic such as *Pseudomonas syringae.* Some strains
295 of *P. fluorescens* can promote the growth of mycelium and ascus opening or other morphological
296 changes in fungi (Citterio *et al* 2001; Cho *et al.,* 2003; Deveau *et al.,* 2007). Certain
297 *Pseudomonas* species have also been reported as bacterial "fungiphiles" in the mycospheres,

298 pointing to their close relationship with fungi (Warmink *et al.,* 2009). Furthermore, *Pedobacter*
299 has been identified as tolaasin detoxifying bacteria from Agaricales (Tsukamoto *et al.,* 2002),
300 and *Chryseobacterium* has been identified in mycospheres of several fungi (Warmink *et al.,*
301 2009). *Herbaspirillum seropedicae* has been reported as nitrogen fixing root associated
302 bacterium (Baldani *et al.,* 1986). Another dominant group, *Burkholderia* is also known to have
303 beneficial interactions with fungi improving the formation of mycorrhiza (Aspray *et al.,* 2006;
304 Frey-Klett *et al.,* 2007) and provide the fungal partner with nutrients in stress conditions
305 (Stopnisek *et al.,* 2016). It is possible that these bacteria play similar roles in epigeous fungal
306 fruiting-bodies; however, more replicated studies are needed to understand the specific functions
307 of these bacteria and to exclude the possibility that they passively colonize fungal fruiting-
308 bodies. Taken together, these data indicate the dominance of symbiotic bacteria in fungal
309 epigeous fruiting-bodies.

310 Despite many similarities, there were some differences between the microbiome of
311 basidiomycetous *I. terrigena* and hypogeous ascomycetous fruiting-bodies (Benucci and Bonito
312 2016; Barbieri *et al.,* 2005, 2007;Quandt *et al.,* 2015). Particularly compared to our dataset
313 where Gamma and Betaproteobacteria were the dominant classes, Alphaproteobacteria and
314 Actinobacteria were the dominant classes in ascomycetous fruiting-bodies. This may be
315 explained by the higher relative abundance of Alphaproteobacteria in soil (Janssen 2006; Fierer
316 *et al.,* 2012; Pent *et al.,* 2017) as well as more intimate association between soil and hypogeous
317 fruiting-bodies, compared to epigeous fruiting-bodies. The differences between bacterial
318 communities of epi- and hypogeous fruiting-bodies may also be explained by different
319 environmental conditions below- and above-ground. In contrast to the microbiome of hypogeous
320 species, which are typically dominated by *Bradyrhizobium* species (Barbieri *et al.,* 2005; 2007;
321 Antony-Babu *et al.,* 2013; Quandt *et al.,* 2015), both *Pseudomonas and Pedobacter* dominated
322 the *I. terrigena microbiome*. Several potato-associated *Pseudomonas* species are able to
323 counteract both plant-pathogenic fungi and plant-parasitic nematode (Krechel *et al.,* 2002), and
324 some *Pedobacter* are associated with soil or plant-pathogenic nematodes (Tian *et al.,* 2011;
325 Baquiran *et al.,* 2013). Thus, it is tempting to suggest that the high abundance of *Pseudomonas*
326 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. S2, S4) 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, and three anonymous reviewers for constructive comments on earlier version of this article. 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

355 Antony-Babu, S., Deveau, A., Nostrand, J.D.V., Zhou, J., Le Tacon, F., Robin, C., Frey-Klett, P.,
356 and Uroz, S. (2013) Black truffle-associated bacterial communities during the development
357 and maturation of *Tuber melanosporum* ascocarps and putative functional roles. *Environ*
358 *Microbiol* **16**:2831–2847.
359 Aspray, T.J., Jones, E.E., Davies, M.W., Shipman, M., and Bending, G.D. (2013) Increased
360 hyphal branching and growth of ectomycorrhizal fungus *Lactarius rufus* by the helper
361 bacterium *Paenibacillus* sp. *Mycorrhiza* **23**:403–410.
362 Aspray, TJ., Jones, EE., Whipps, J.M., and Bending, G.D. (2006) Importance of mycorrhization
363 helper bacteria cell density and metabolite localization for the *Pinus sylvestris-Lactarius rufus*
364 symbiosis. FEM*S Microbiol Ecol* **56**:25–33.
365 Baldani, J.I., Baldani, V.L.D., Seldin, L., and Döbereiner, J. (1986) Characterization of
366 *Herbaspirillum seropedicae* gen. nov., sp. nov., a root-associated nitrogen-fixing bacterium.
367 *Int J Syst Evol Microbiol* **36**:86–93.
368 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M.,
369 Nikolenko, S.I., Pham, S., and Prjibelski, A.D. (2012) SPAdes: a new genome assembly
370 algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**:455–477.
371 Baquiran, J.P., Thater, B., Sedky, S., De Ley, P., Crowley, D., and Orwin, P.M. (2013) Culture-
372 Independent investigation of the microbiome associated with the nematode *Acrobeloides*
373 *maximus*. *PLOS ONE* **8**:e67425.
374 Barbieri, E., Bertini, L., Rossi, I., Ceccaroli, P., Saltarelli, R., Guidi, C., Zambonelli, A., and
375 Stocchi, V. (2005) New evidence for bacterial diversity in the ascoma of the ectomycorrhizal
376 fungus *Tuber borchii* Vittad. *FEMS Microbiol Let* **247**:23–35.
377 Barbieri, E., Guidi, C., Bertaux, J., Frey-Klett, P., Garbaye, J., Ceccaroli, P., Saltarelli, R.,
378 Zambonelli, A., and Stocchi, V. (2007) Occurrence and diversity of bacterial communities in
379 Tuber magnatum during truffle maturation. *Environ Microbiol* 9:2234–2246.
380 Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., Wit, P.,
381 Sánchez-García, M., Ebersberger, I., and Sousa, F. (2013) Improved software detection and
382 extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for
383 analysis of environmental sequencing data. *Methods Ecol Evol* 4:914–919.

384 Benucci, G.M.N., and Bonito, G.M. (2016) The truffle microbiome: species and geography
385 effects on bacteria associated with fruiting bodies of hypogeous Pezizales. *Microbiol Ecol*
386 **72**:4–8.
387 Bergmann, G.T., Bates, S.T., Eilers, K.G., Lauber, C.L., Caporaso, J.G., Walters, W.A., Knight,
388 R., and Fierer, N. (2011) The under-recognized dominance of Verrucomicrobia in soil
389 bacterial communities. *Soil Biol Biochem* **43**:1450–1455.
390 Chen, S., Qiu, C., Huang, T., Zhou, W., Qi, Y., Gao, Y., Shen, J., and Qiu, L. (2013) Effect of 1-
391 aminocyclopropane-1-carboxylic acid deaminase producing bacteria on the hyphal growth
392 and primordium initation of Agaricus bisporus. *Fung Ecol* **6**:110–118.
393 Cho, Y.S., Kim, J.S., Crowley, D.E., and Cho, B.G. (2003) Growth promotion of the edible
394 fungus *Pleurotus ostreatus* by fluorescent pseudomonads. *FEMS Microbio Lett* **218**:271–276.
395 Citterio, B., Malatesta, M., Battistelli, S., Marcheggiani, F., Baffone, W., Saltarelli, R., Stocchi,
396 V., and Gazzanelli, G. 2001. Possible involvement of *Pseudomonas fluorescens* and
397 Bacillaceae in structural modifications of Tuber borchii fruit bodies. *Can J Microbiol* **47**:
398 264–268.
399 Dahm, H., Wrótniak, W., Strzelczyk, E., and Bednarska, E. (2005) Diversity of Culturable
400 Bacteria associated with fruiting bodies of ectomycorrhizal fungi. *Phytopathol Pol* **38**: 51–62.
401 Danell, E., Alström, S., and Ternström, A. (1993) *Pseudomonas fluorescens* in association with
402 fruitbodies of the ectomycorrhizal mushroom *Cantharellus cibarius*. *Mycol Res* **97**:1148–
403 1152.
404 Deveau, A., Palin, B., Delaruelle, C., Peter, M., Kohler, A., Pierrat, J.C., Sarniguet, A., Garbaye,
405 J., Martin, F., and Frey-Klett, P. (2007) The mycorrhiza helper Pseudomonas fluorescens
406 BBc6R8 has a specific priming effect on the growth, morphology and gene expression of the
407 ectomycorrhizal fungus *Laccaria bicolor* S238N. *New Phytol* **175**:743–755.
408 Dominguez, JA., Martin, A., Anriquez, A., and Albanesi, A. (2012) The combined effects of
409 *Pseudomonas fluorescens* and *Tuber melanosporum* on the quality of *Pinus halepensis*
410 seedlings. *Mycorrhiza* **22**:429–436.
411 Doré, J., Perraud, M., Dieryckx, C., Kohler, A., Morin, E., Henrissat, B., Lindquist, E.,
412 Zimmermann, SD., Girard, V., and Kuo, A. (2015) Comparative genomics, proteomics and
413 transcriptomics give new insight into the exoproteome of the basidiomycete *Hebeloma*

414 *cylindrosporum* and its involvement in ectomycorrhizal symbiosis. *New Phytol* **208**:1169–
415 1187.
416 Eilers, K.G., Lauber, C.L., Knight, R., and Fierer, N. (2010) Shifts in bacterial community
417 structure associated with inputs of low molecular weight carbon compounds to soil. *Soil Biol*
418 *Biochem* **42**:896–903.
419 Fierer, N., Left, J.W., Adams, B.J., Nielsen, U.N., Bates, S.T., Lauber, C.L., Owens, S., Gilbert,
420 J.A., Wall, D.H., and Caporaso, J.G. (2012) Cross-biome metagenomic analyses of soil
421 microbial communities and their functional attributes. *Proc Natl Acad Sci USA* **109**:21390–
422 21395.
423 Floudas, D., Binder, M., Riley, R., Barry, K., Blanchette, R.A., Henrissat, B., Martinez, A.T., et
424 al. (2012) The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31
425 fungal genomes. *Science* **336**:1715–1719.
426 Frey-Klett, P., Garbaye, J., and Tarkka, M. (2007) The mycorrhiza helper bacteria revisited. *New*
427 *Phytol* **176**:22–36.
428 Haas, D., and Defago, G. (2005) Biological control of soil-borne pathogens by fluorescent
429 pseudomonads. *Nat Rev Microbiol* **3**:307-19.
430 Halsey, J.A., e Silva, M.D.C.P., and Andreote, F.D. (2016) Bacterial selection by mycospheres
431 of Atlantic Rainforest mushrooms. *Antonie Leeuwenhoek* **109**:1353–1365.
432 Han, M.V., Thomas, G.W., Lugo-Martinez, J., and Hahn, M.W. (2013) Estimating gene gain and
433 loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol*
434 *Evol* **30**:1987–1997.
435 Hibbett, D.S., Gilbert, L-B., and Donoghue, M.J. (2000) Evolutionary instability of
436 ectomycorrhizal symbioses in basidiomycetes. *Nature* **407**:506–508.
437 Hibbett, D., Grimaldi, D., and Donoghue M. (1997) Fossil mushrooms from Miocene and
438 Cretaceous ambers and the evolution of Homobasidiomycetes. *Am J Bot* **84**:981.
439 Holt, C., and Yandell, M. (2011) MAKER2: an annotation pipeline and genome-database
440 management tool for second-generation genome projects. *BMC Bioinform*. **12**:491.
441 Huerta-Cepas, J., Forslund, K., Pedro Coelho, L., Szklarczyk, D., Juhl Jensen, L., von Mering, C.
442 and Bork, P. (2017) Fast genome-wide functional annotation through orthology assignment by
443 eggNOG-mapper *Mol Bio Evol* **34**:2115-2122.

444 Hyatt, D., Chen, G-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010)
445 Prodigal: prokaryotic gene recognition and translation initiation site *identification. BMC*
446 *Bioinform* **11**:119.
447 Janssen, P.H. (2006) Identifying the Dominant Soil Bacterial Taxa in Libraries of 16S rRNA and
448 16S rRNA Genes. *Appl Environ Microbiol* **72**:1719–1728.
449 Jones, P., Binns, D., Chang, H-Y., Fraser, M., Li ,W., McAnulla, C., McWilliam, H., Maslen, J.,
450 Mitchell, A., and Nuka, G. (2014) InterProScan 5: genome-scale protein function
451 classification. *Bioinformatics* **30**:1236–1240.
452 Kanehisa M., Sato Y., and Morishima K. (2016) BlastKOALA and GhostKOALA: KEGG tools
453 for functional characterization of genome and metagenome sequences. *J Mol Biol* **428**:726–
454 731.
455 Katoh, K., and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7:
456 improvements in performance and usability. *Mol Biol Evol* **30**:772–780.
457 Kohler, A., Kuo, A., Nagy, L.G., Morin, E., Barry, K.W., Buscot, F., Canbäck, B., Choi, C.,
458 Cichocki, N., and Clum, A. (2015) Convergent losses of decay mechanisms and rapid
459 turnover of symbiosis genes in mycorrhizal mutualists. *Nat Genet* **47**:410–415.
460 Krechel, A., Faupel, A., Hallmann, J., Ulrich, A., and Berg, G. (2002) Potato-associated bacteria
461 and their antagonistic potential towards plant-pathogenic fungi and the plant-parasitic
462 nematode *Meloidogyne incognita* (Kofoid & White) Chitwood. *Can J Microbiol* 48:772–786
463 Kõljalg, U., Nilsson, RH., Abarenkov, K., Tedersoo, L., Taylor, A.F., Bahram, M., *et al.* (2013)
464 Towards a unified paradigm for sequence-based identification of fungi. *Mol Ecol* **21**:5271–
465 7.
466 Kumari, D., Reddy, M.S., and Upadhyay, R.C. (2013) Diversity of cultivable bacteria associated
467 with fruiting bodies of wild Himalayan *Cantharellus* spp. *Ann Microbiol* **63**:845-853.
468 Lackner, G., Partida-Martinez, L.P., and Hertweck, C. (2009) Endofungal bacteria as producers
469 of mycotoxins. *Trends Microbiol* **17**:570–576.
470 Laetsch, D.R., and Blaxter, M.L. (2017) BlobTools: Interrogation of genome assemblies.
471 *F1000Research* **6**:1287.
472 Lagesen, K., Hallin, P., Rødland, E.A., Stærfeldt, H.H., Rognes, T. and Ussery, D.W. (2007)
473 RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucl Acid Res* **35**:3100-
474 3108.

475 Langmead, B., and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature*
476 *Methods* **9**:357–359.
477 Larsen, T.S., and Krogh, A. (2003) EasyGene–a prokaryotic gene finder that ranks ORFs by
478 statistical significance. *BMC Bioinform* 4:21.
479 Li L., Stoeckert, C.J., and Roos, D.S. (2003) OrthoMCL: identification of ortholog groups for
480 eukaryotic genomes. *Genome Res* **13**:2178–2189.
481 Martin, F., Kohler, A., Murat, C., Balestrini, R., Coutinho, P.M., Jaillon, O., Montanini, B.,
482 Morin, E., Noel, B., and Percudani, R. (2010) Périgord black truffle genome uncovers
483 evolutionary origins and mechanisms of symbiosis. *Nature* **464**:1033–1038.
484 Matheny, P.B., Aime, M.C., Bougher N.L., Buyck, B., Desjardin, D.E., Horak, E., Kropp, B.R.,
485 Lodge, D.J., Soytong, K., Trappe, J.M., and Hibbett, D.S. (2009) Out of the Palaeotropics?
486 Historical biogeography and diversification of the cosmopolitan ectomycorrhizal mushroom
487 family Inocybaceae. *J Biogeogr* **36**:577–92.
488 Matheny, P.B., Curtis, J.M., Hofstetter, V., Aime, M.C., Moncalvo, J.M., Ge, Z.W., Yang, Z.L.,
489 Slot, J.C., Ammirati, J.F., Baroni, T.J. and Bougher, N.L. (2006) Major clades of Agaricales:
490 a multilocus phylogenetic overview. *Mycologia* **98**: 982-995.
491 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella,
492 K., Altshuler, D., Gabriel, S., and Daly, M. (2010) The Genome Analysis Toolkit: a
493 MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*
494 **20**:1297–1303.
495 Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, EM., Kubal, M., Paczian, T.,
496 Rodriguez, A., Stevens, R., and Wilke, A. (2008). The metagenomics RAST server–a public
497 resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC*
498 *Bioinform* **9**:386.
499 Minh, B,Q., Nguyen, M.A.T., and von Haeseler, A. (2013) Ultrafast approximation for
500 phylogenetic bootstrap. *Mol Biol Evol* **30**:1188–1195.
501 Mirarab, S., and Warnow, T. (2015) ASTRAL-II: coalescent-based species tree estimation with
502 many hundreds of taxa and thousands of genes. *Bioinformatics* **31**: i44–i52.
503 Mu, L.L., Yun, Y.B., Park, S.J., Cha, J.S., and Kim, Y.K. (2015) Various pathogenic
504 Pseudomonas strains that cause brown blotch disease in cultivated mushrooms. *J Appl Biol*
505 *Chem* **58**:349–354.

506 Munsch, P., Alatossava, T., Marttinen, N., Meyer, J.M., Christen, R., and Garden, L. (2002)
507 *Pseudomonas costantinii* sp. nov., another causal agent of brown blotch disease, isolated from
508 cultivated mushroom sporophores in Finland. *Int J Syst Evol Microbiol* **52**:1973–1983.
509 Nazir, R., Warmink, J.A., Boersma, H., and van Elsas, J.D. (2010a) Mechanisms that promote
510 bacterial fitness in fungal-affected soil microhabitats. *FEMS Microbiol Ecol* **71**:169–185
511 Nazir, R., Boersma, F.G.H., Warmink, J.A., and van Elsas, J.D. (2010b) *Lyphyllum* sp. strain
512 Karsten alleviates pH pressure in acid soil and enhances the survival of *Variovorax paradoxus*
513 HB44 and other bacteria in the mycosphere. *Soil Biol Biochem* **42**:2146–2152
514 Olson, Å., Aerts, A., Asiegbu, F., Belbahri, L., Bouzid, O., Broberg, A., Canbäck, B., Coutinho,
515 PM., Cullen, D., and Dalman, K. (2012) Insight into trade-off between wood decay and
516 parasitism from the genome of a fungal forest pathogen. *New Phytol* **194**:1001–1013.
517 Parra, G., Bradnam, K., and Korf, I. (2007) CEGMA: a pipeline to accurately annotate core
518 genes in eukaryotic genomes. *Bioinformatics* **23**:1061–1067.
519 Quandt, CA., Kohler, A., Hesse, C., Sharpton, T., Martin, F., and Spatafora, J.W. (2015)
520 Metagenome sequence of Elaphomycesgranulatus from sporocarp tissue reveals Ascomycota
521 ectomycorrhizal fingerprints of genome expansion and a Proteobacteria rich microbiome.
522 *Environ Microbiol* **17**:2952–2968.
523 Pent, M., Põldmaa, K., and Bahram, M. (2017) Bacterial communities in boreal forest
524 mushrooms are shaped both by soil parameters and host identity. *Front Microbiol* **8**:836.
525 Riedlinger, J., Schrey, S.D., Tarkka, M.T., Hampp, R., Kapur, M., and Fiedler, H-P. (2006)
526 Auxofuran, a novel metabolite that stimulates the growth of fly agaric, is produced by the
527 mycorrhiza helper bacterium Streptomyces strain AcH 505. *Appl Environ Microbiol* **72**:3550–
528 3557.
529 Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S.R., Wilkie, A.O., McVean, G., Lunter,
530 G., and WGS500 Consortium. (2014) Integrating mapping-, assembly-and haplotype-based
531 approaches for calling variants in clinical sequencing applications. *Nat Genet* **46**:912–918.
532 Smit, A., and Hubley, R. (2011) RepeatModeler-1.0. 5. Institute of Systems Biology.
533 Smit A.F, Hubley R, Green P. 2010. RepeatMasker Open-3.0. . *URL:* http://www.repeatmasker.org/
534
535 Smith SE., and Read DJ. (2010) Mycorrhizal symbiosis. 3rd edn. London, UK: Academic Press.

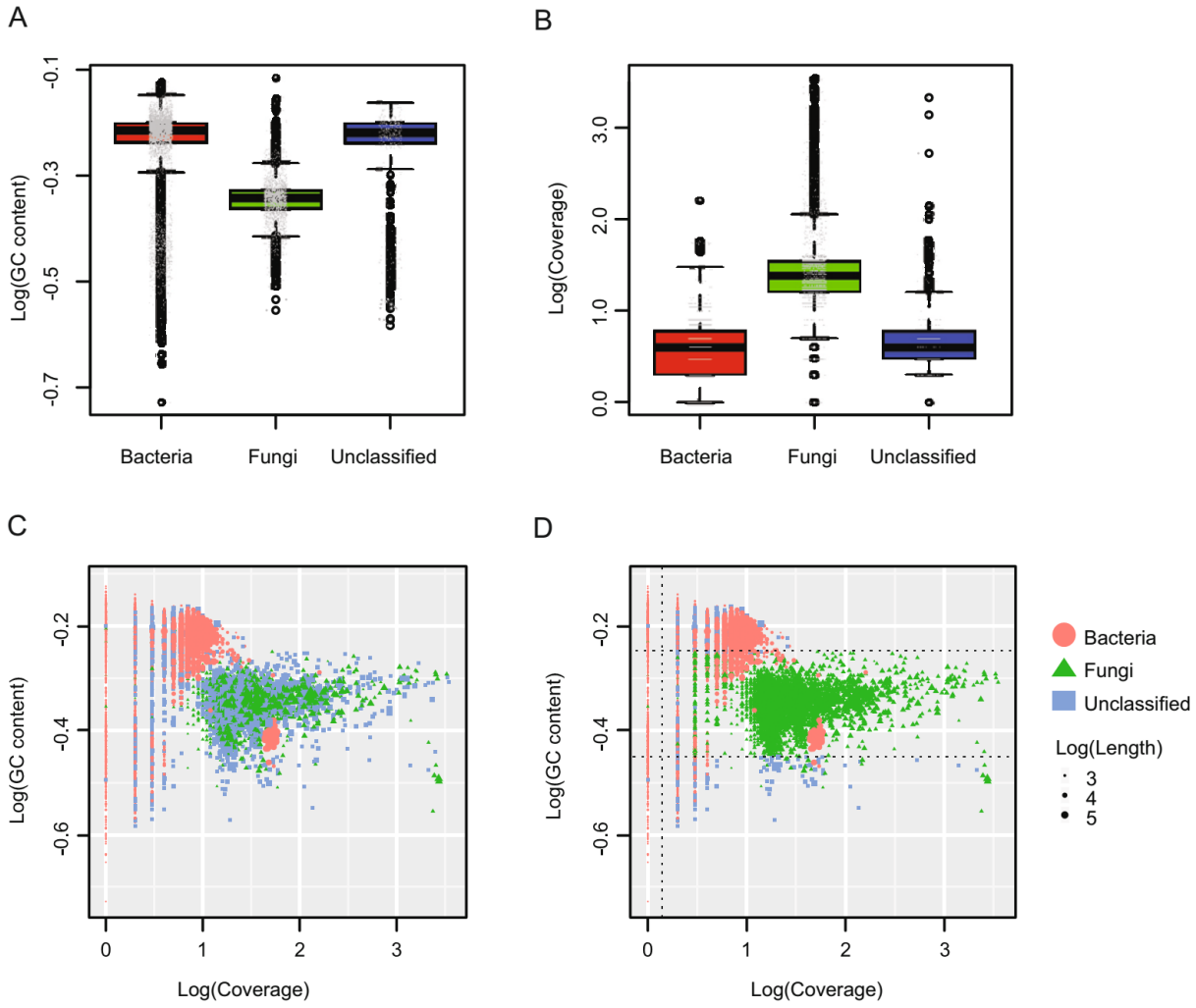
536 Spanu, P.D., Abbott, J.C., Amselem, J., Burgis, T.A., Soanes, D.M., Stuber, K., et al. (2010)
537 Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme
538 parasitism. *Science* **330**:1543–1546.
539 Splivallo, R., Deveau, A., Valdez, N., Kirchhoff, N., Frey-Klett, P., and Karlovsky, P. (2015)
540 Bacteria associated with truffle-fruiting bodies contribute to truffle aroma. *Environ Microbiol*
541 **17**:2647–2660.
542 Stamatakis A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of
543 large phylogenies. *Bioinformatics* **30**:1312–1313.
544 Stopnisek, N., Zühlke, D., Carlier, A., Barberán, A., Fierer, N., Becher, D., Riedel, K., Eberl L.,
545 and Weisskopf, L. (2016) Molecular mechanisms underlying the close association between
546 soil Burkholderia and fungi. *ISME* **10**:253–264.
547 Stubblefield, S,P., Taylor, T,N., and Beck, C.B. (1985) Studies of paleozoic fungi. IV. Wood-
548 decaying fungi in Callixylon newberryi from the upper Devonian. *Am J Bot* 1765-1774.
549 Talavera, G., and Castresana, J. (2007) Improvement of phylogenies after removing divergent
550 and ambiguously aligned blocks from protein sequence alignments. *Syst Boil* **56**:564–577.
551 Tedersoo, L., May T.W., and Smith, M.E. (2010) Ectomycorrhizal lifestyle in fungi: global
552 diversity, distribution, and evolution of phylogenetic lineages. *Mycorrhiza* **20**:217–263.
553 Thorne, J.L., Kishino, H., and Painter, I.S. (1998) Estimating the rate of evolution of the rate of
554 molecular evolution. *Mol Biol Evol* **15**:1647–1657.
555 Tian, X., Cheng, X., Mao, Z., Chen, G., Yang, J., and Xie, B. (2011) Composition of Bacterial
556 communities associated with plant-parasitic nematode *Bursaphelenchus mucronatus*. *Curr*
557 *Microbiol* **62**:117–125.
558 Torres-Cortés, G., Ghignone, S., Bonfante, P., and Schüßler, A. (2015) Mosaic genome of
559 endobacteria in arbuscular mycorrhizal fungi: Transkingdom gene transfer in an ancient
560 mycoplasma-fungus association. *Proc Natl Acad Sci USA* **112**:7785–7790.
561 Tsukamoto, T., Murata, H., and Shirata, A. (2002) Identification of non-pseudomonad bacteria
562 from fruit bodies of wild agaricales fungi that detoxify tolaasin produced by *Pseudomonas*
563 *tolaasii*. *Biosci Biotech Biochem* **66**:2201–2208.
564 UniProt Consortium (2014) UniProt: a hub for protein information. *Nucl Acid Res* gku989.

565 Uroz. S., Oger. P., Morin. E., and Frey-Klett. P. (2012) Distinct ectomycorrhizospheres share
566 similar bacterial communities as revealed by pyrosequencing-based analysis of 16S rRNA
567 genes. *Appl Environ Microbiol* **78**:3020–3024.
568 Veldre. V., Abarenkov. K., Bahram. M., Martos. F., Selosse. M-A., Tamm. H., Kõljalg. U., and
569 Tedersoo. L. (2013) Evolution of nutritional modes of Ceratobasidiaceae (Cantharellales,
570 Basidiomycota) as revealed from publicly available ITS sequences. *Fung Ecol* **6**:256–268.
571 Vicente. CSL., Nascimento. F., Espada. M., Barbosa. P., Mota, M., Glick, B.R., and Oliveira, S.
572 (2012) Characterization of bacteria associated with pinewood nematode *Bursaphelenchus*
573 *xylophilus*. *PLOS ONE* **7**: e46661.
574 Wang, H.C., Minh, B.Q., Susko, E., and Roger, A.J. (2017) Modeling Site Heterogeneity with
575 Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. *Syst*
576 *Bio* doi: 10.1093/sysbio/syx068.
577 Warmink, JA., Nazir, R., and van Elsas, J.D. (2009) Universal and species-specific bacterial
578 "fungiphiles" in the mycospheres of different basidiomycetous fungi. *Environ Microbiol*
579 **11**:300–312.
580 Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012) dbCAN: a web resource for
581 automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **40**:W445–451.
582
583
584
585
586
587
588
589
590
591
592
593
594

595

596 Figures

597

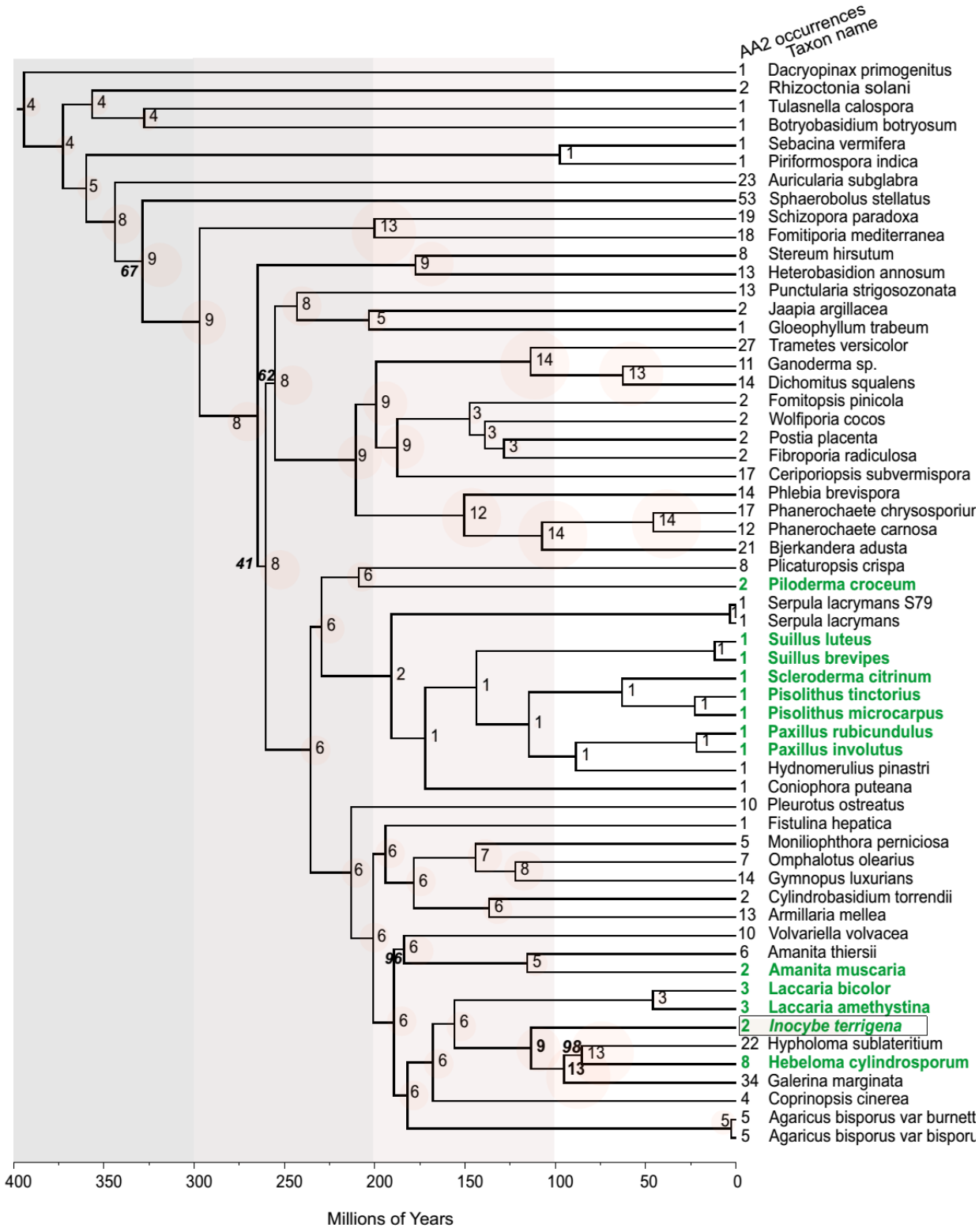


598

599 **Figure 1**. Bacterial and fungal assemblies can be separated based on GC content and coverage.
600 (A-B) The boxplots of GC content (A) and median coverage (B) for bacterial, fungal contigs
601 identified based on Blast searches. (C) The scatterplot of GC content as a function of median
602 coverage of contigs. (D) Same as C however unclassified contigs with median coverage >1 and

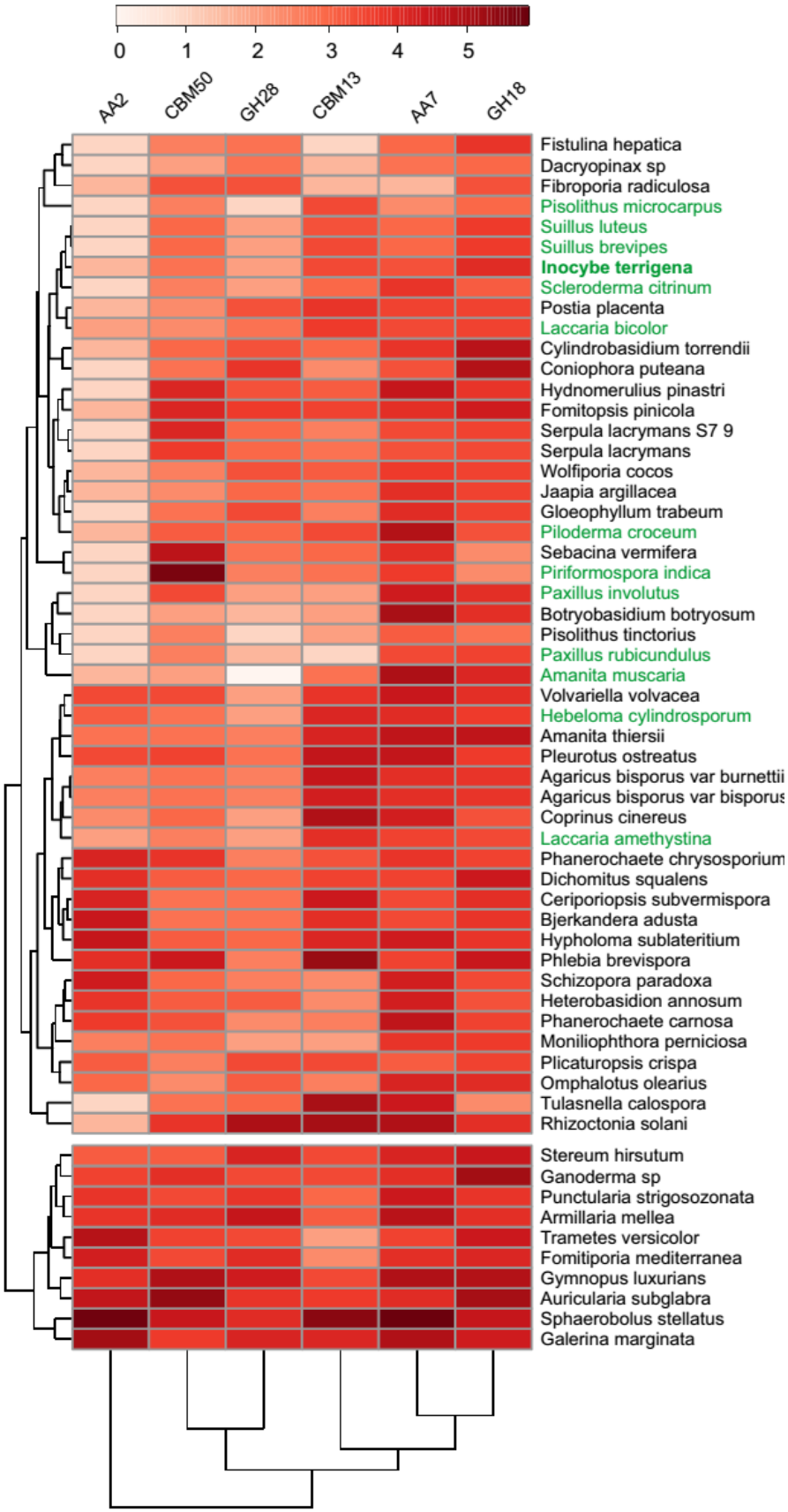
603 GC content >0.35 and <0.56 were classified as fungi. In C and D, unclassified contigs of <1000
604 bp length have been excluded.

605


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 chrysosporiur
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 Pleurotus ostreatus
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 burnett
5 Agaricus bisporus var bisporu
400 350 300 250 200 150 100 50 0
Millions of Years

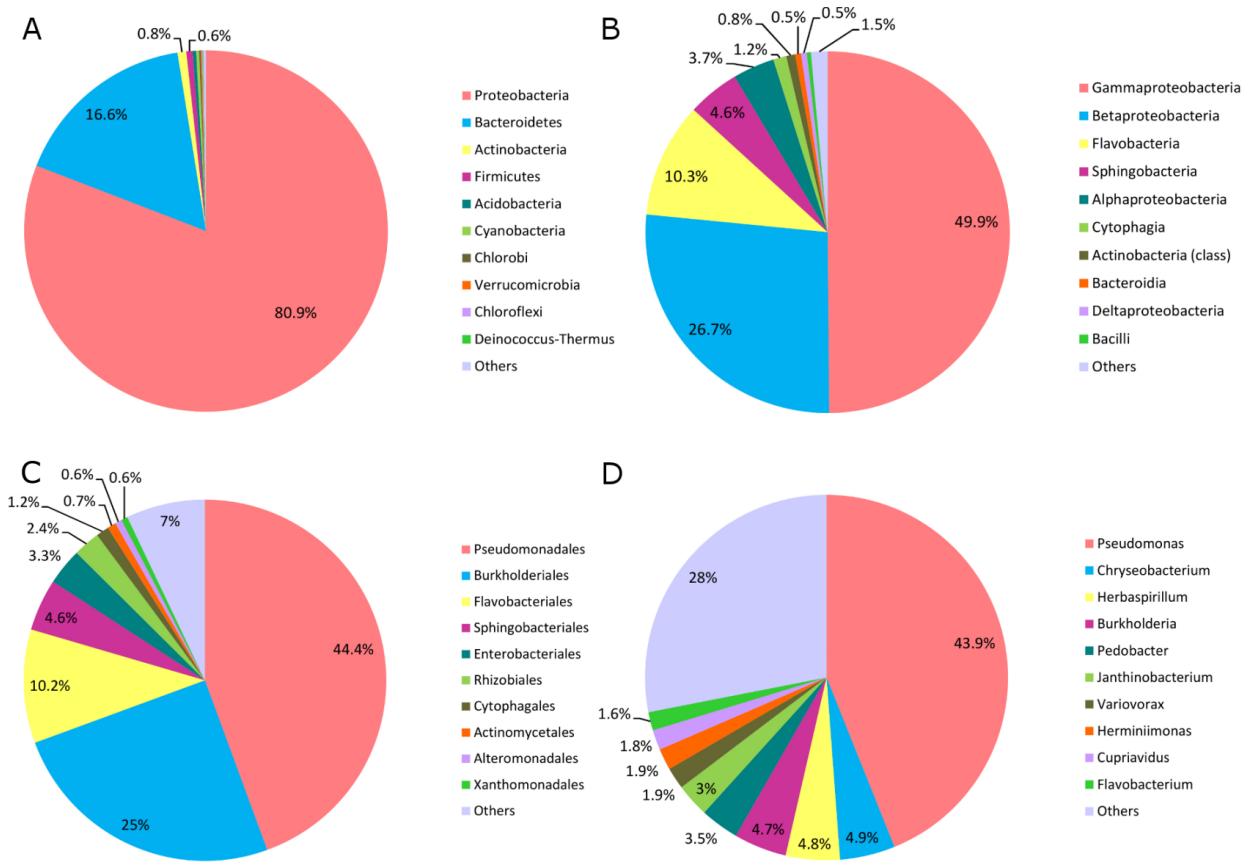
606

607 **Figure 2**. Phylogenetic tree showing the phylogenetic relationship of *Inocybe terrigena* and
608 other published ectomycorrhizal and saprotrophic fungi. All clades have a support value of 100%
609 except those that are indicted in italic text. Non-italic node labels and numbers next to taxon
610 names represent the number of AA2 gene occurrences for each parent node and taxon, indicating
611 gain/loss of AA2 genes in each taxon. Note that *I. terrigena* belongs to a clade which has gained
612 three AA2 genes from its parent node, but it has lost seven AA2 genes compared to other
613 members of this clade. Numbers on the tree nodes and edges are estimated (based on CAFE) and
614 observed copy numbers of AA2.


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
Scleroderma citrinum
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

615

616 **Figure 3**. Heatmap of six CAZyme families that showed significant expansion or contraction
617 across 59 analysed genomes in this study. The scale shows the copy numbers (Log) of CAZyme genes
618 in each family.



619

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

625