

***New Phytologist* Supporting Information**

Article title: RNA-based analyses reveal fungal communities structured by a senescence gradient in the moss *Dicranum scoparium* and the presence of putative multi-trophic fungi

Authors: Ko-Hsuan Chen^{1*}, Hui-Ling Liao^{1,2}, A. Elizabeth Arnold³, Gregory Bonito⁴, François Lutzoni¹

Article acceptance date: 7 February 2018

Supporting Information Methods S1-S7

Supporting Information Figures S1-S10

Supporting Information Methods S1-S7.

Methods S1. RNAseq library preparation

Prior to RNAseq library preparation, the quality of the isolated RNA was evaluated with both Qubit Fluorometric Quantitation (Thermo Fisher Scientific, Waltham, MA, USA) and Bioanalyzer analysis (Agilent Technologies, Inc., Santa Clara, CA, USA) (Supplementary Table S4). The cDNA library from the metatranscriptome was prepared with the Illumina TruSeq Stranded mRNA library Prep kit (Illumina Inc., San Diego, CA, USA) following manufacturer's protocol. The nine barcoded cDNA libraries (three layers by three mats) were pooled together and sequenced using one Illumina HiSeq 2500 lane (125 bp paired end) at the Duke Center for Genomic and Computational Biology.

Methods S2. nrLSU cDNA library preparation for amplicon sequencing

Amplicon sequencing was not performed for the bryophyte mat sampled at microsite 1 because insufficient RNA remained after the RNAseq library preparation of that collection. To prepare cDNA libraries for amplicon sequencing, the RNA extractions were first treated with RNase-free DNase (Qiagen, Hilden, Germany). The dNTPs and primers were diluted in Diethylpyrocarbonate (DEPC)-treated water. Eight microliters of RNA (RNA concentration $35.98 \pm \text{SD } 13.31 \text{ ng}/\mu\text{l}$) were used for each sample. The RNA was reverse transcribed to a cDNA library with the SuperScript® II Reverse Transcriptase kit (Sigma) using the LR3 primer following the manufacturer's instruction, yielding cDNAs with concentrations of $2160 \pm \text{SD } 207.75 \text{ (ng}/\mu\text{l)}$ across the six samples. When preparing MiSeq libraries, each 25 μl reaction consisted of 0.5 unit Qiagen HotStart Plus polymerase (Qiagen, Hilden, Germany), 1X Qiagen PCR buffer, 0.16 mM (each) deoxynucleoside triphosphates (dNTPs), 0.75 mM MgCl_2 , 0.6 mM forward and reverse primers, 1 mg/mL bovine serum albumin (BSA). For the first PCR round 1 μl of DNA was used as DNA template. For the second PCR round, 2 μl of PCR product from round 1 was used for

as DNA template. In the third PCR round, 10µl of PCR product was used as DNA template. Thermocycler settings were 5 min at 95 °C, then 10 cycles of 95 °C for 1 min, 52 °C for 2 min, and 72 °C for 2 min extension, with a final extension for 10 min at 72 °C. Prepared libraries were normalized, were pooled with other libraries (239 libraries in total), and unincorporated primers, dNTPs and primer dimers were removed by two successive rounds of cleaning using the Agencourt AMPure purification system (Beckman Coulter, Danvers, MA, USA). Pooled amplicon libraries were sequenced as part of one Illumina MiSeq run (300 bp paired end).

Methods S3. Additional information on RNAseq data quality checking and processing

Trimmomatic (Bolger et al., 2014) was used to perform quality filtering and adaptor trimming (Supplementary Method S5). Bases were cut when the average quality of the 4-base sliding window drop below Phred=15. FastQC in the Galaxy portal was used to examine the quality of the reads (Blankenberg et al., 2010). All metatranscriptome mapping procedures were performed with Bowtie2 with default parameters (default settings see Bowtie2 manual v2.2.6 [<http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml>]). Samtool (Li et al., 2009) was used to generate mapping reports and map sequences to the targeted reference.

Methods S4. MiSeq amplicon sequence processing

We first trimmed the adaptors using cutadapt with 10% mismatch allowance (Martin, 2011) followed by quality filtering with USEARCH (Edgar, 2013). All reads were truncated to 150 bp with the maximum error rate allowance of 0.25. Prior to clustering reads into operational taxonomic units (OTU), all samples were pooled together and singletons were removed. OTUs were clustered based on 97% similarity using UPARSE. *De novo* chimera filtering and representative sequence selection were part of the OTU clustering process. OTU tables were generated by mapping original reads (before the removal of singletons) to the representative sequences with 97% similarity (see Supplementary Method S5 for scripts).

Methods S5. Scripts for bioinformatics

The programs or resources used are in bold and italic. The purposes of the scripts are in bold. Scripts are preceded by the pound (#) sign.

GenBank (NCBI) search

Terms for NCBI database search when constructing primary database

(fungi[Organism]) AND (5S or 18S or "Small Subunit ribosomal" or "internal transcribed" or LSU or SSU or ITS or 5.8s or 28s or 25s or 26s or "Large Subunit ribosomal") NOT (supercont OR shotgun OR supercontig OR contig OR mRNA OR beta-tubulin OR hypothetical OR uncultured OR clone OR snoRNA OR Patent OR ORF OR -like OR NTS)

Trimmomatic

For quality check and trimming of RNAseq reads

```
#java -jar trimmomatic-0.32.jar PE example_R1.gz example_R2.gz  
1T_forward_paired.fq.gz 1T_forward_unpaired.fq.gz 1T_reverse_paired.fq.gz  
1T_reverse_unpaired.fq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:28:10 LEADING:3  
TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36
```

cutadapt

For amplicon sequencing LR0R adaptor removal:

```
# ./cutadapt -g  
GCCTCCCTCGCGCCATCAGAGATGTGTATAAGAGACAGNNNNNNNNNNNGAACCCGCT  
GAACTTAAGC -o example_trimmed.fastq example.fastq --untrimmed-output  
example_untrimmed.fastq
```

For amplicon sequencing LR3 adaptor removal:

```
# ./cutadapt -g  
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNCACCGTGTTTCA
```



```
AGACGGG -o example_trimmed.fastq example.fastq --untrimmed-output  
example_untrimmed.fastq
```

For metatranscriptome primary library preparation-checking LR0R+LR15

```
#!/cutadapt -g GTCCGAGTTGTAATTTA -e 0.15 -o  
contain_LR15_GenBank_ref_11302015.fasta --no-trim --overlap=12 --minimum-  
length 100 --untrimmed-output untrimmedLR15_LROR_GenBank_ref_11302015.fas  
GenBank_ref_11302015.fasta  
#!/cutadapt -g ACCCGCTGAACTTAAGC -e 0.15 -o  
contain_LR15+LROR_GenBank_ref_11302015.fasta --overlap=12 --minimum-length  
100 --untrimmed-output untrimmed_LROR_GenBank_ref_11302015.fasta  
contain_LR15_GenBank_ref_11302015.fasta
```

For metatranscriptome primary library preparation-checking ITS4+LR3

```
#!/cutadapt -a GTCTTGAAACACGGACC -e 0.15 -o  
contain_LR3_GenBank_ref_11302015.fasta --no-trim --overlap=12 --minimum-  
length 100 --untrimmed-output untrimmedLR3_LROR_GenBank_ref_11302015.fas  
GenBank_ref_11302015.fasta  
#!/cutadapt -g ACCCGCTGAACTTAAGC -e 0.15 -o  
contain_LR3+LROR_GenBank_ref_11302015.fasta --overlap=12 --minimum-length  
100 --untrimmed-output untrimmed_LROR_GenBank_ref_11302015.fasta  
contain_LR3_GenBank_ref_11302015.fasta
```

Check sequences with LR3, then trimmed with LR0R

```
#!/cutadapt -a GTCTTGAAACACGGACC -e 0.15 --overlap=12 -o  
LSU_DNA_fungi_wLR3.fasta --no-trim --untrimmed-output no_LR3.fasta --minimum-  
length 100 LSU_DNA_fungi.fasta  
  
#!/cutadapt -g ACCCGCTGAACTTAAGC -e 0.15 --overlap=12 -o  
LSU_DNA_fungi_wLR3LROR.fasta --untrimmed-output wLR3_noLROR.fasta --  
minimum-length 100 LSU_DNA_fungi_wLR3.fasta
```

UPARSE pipeline

Quality filter and truncation

```
# usearch8 -fastq_filter trimmed_H242SR_LSU_R2.fastq -fastq_trunclen 150 -  
fastq_maxee 0.25 -fastaout H242SR_LR3_L150E025.fasta &>  
H242SR_LR3_L150E025_filter.log
```

Add barcode to reads

```
# sed "-es/^>\(.*\)/>\1;barcodelabel=3B;/" < 3B_ITS4_L150E025.fasta > 3B_ITS4.fa
```

Dereplicate reads and add size (number of the same read) to the reads

```
#usearch8 -derep_fulllength H3LSU_pine.fa -fastaout H3LSU_pine_derep.fas -  
sizeout  
usearch8 -sortbysize H3LSU_pine_derep.fas -minsize 2 -fastaout  
H3LSU_pine_derep_mc2.fasta
```

Remove singletons

```
#usearch8 -sortbysize H3LSU_pine_derep.fas -minsize 2 -fastaout  
H3LSU_pine_derep_mc2.fasta
```

OTU clustering, representative sequences selection and *De novo* chimera check

```
#usearch8 -cluster_otus LR3_L150E025.derep.mc2.fasta -otus  
LR3_L150E025.derep.mc2.repset.fasta &> LR3_L150E025.derep.mc2.otu97.log
```

Numbering OTUs

```
#python fasta_number.py H3LSU_pine_derep_mc2_represt.fas OTU_>  
H3LSU_pine_derep_mc2_represt_label.fasta
```

mapping original reads to representative sequences of OTUs

```
#usearch8 -usearch_global H3LSU_pine.fa -db  
H3LSU_pine_derep_mc2_represt_label.fasta -strand plus -id 0.97 -uc  
H3LSU_pine_derep_mc2_represt_label_map.uc -threads 24
```

QIIME

Normalize read count

```
#normalize_table.py -i LSU_L150E025.derep.mc2.repset_fungi_label_map_sort.biom  
-z --DESeq_negatives_to_zero -a DESeq2 -o  
LSU_L150E025.derep.mc2.repset_fungi_label_map_sort_DESeq2_normalized_otu_table.biom
```

Calculate beta diversity (Bray-Curtis index)

```
#beta_diversity.py -i allcount_forQiime.biom -m bray_curtis -o bray_curtis.txt
```

Generating taxa summary and barplot

```
#summarize_taxa_through_plots.py -o qiime_1 -i allcount_forQiime_tax.biom -m  
all_GB+customized_silva_wLRORLR3_1208/velvet60/map.tsv
```

Adding taxonomy to biom data

```
#biom add-metadata --sc-separated taxonomy --observation-header  
OTUID,taxonomy --observation-metadata-fp  
H3LSU_pine_derep_mc2_represt_label_tax_assignments.txt -i  
H3LSU_pine_derep_mc2_represt_label_map-table.biom -o  
H3LSU_pine_derep_mc2_represt_label_map-table_rdp.biom
```

NMDS

```
#nmads.py -i bray_curtis_moss_culture_list_otutable_deseq2.txt -d 2 -o NMDS
```

Turnover and nestedness measurements (in R package “betapart”)

```
#beta.multi(X, index.family="sorensen")
```

Methods S6. Primary and secondary (MA, MP) reference database constructions

To detect fungal activity, ribosomal RNA abundances were examined using databases prepared with LR0R (Vilgalys unpublished [<http://www.botany.duke.edu/fungi/mycolab>]) and LR3 (Vilgalys and Hester, 1990) primers. The LR0R (forward primer) and LR3 (reverse primer) are commonly used for targeting D1 and D2 regions of the nrLSU, respectively (Hinrikson et al., 2005). The LR0R/LR3 databases were constructed through several steps shown in Supplementary Fig S2a, S3a. First, ribosomal RNA data were combined from three sources: 1) GenBank (accessed on 30 November 2015 [<http://www.ncbi.nlm.nih.gov/>]); 2) Silva database (accessed on 14 October 2014 [<https://www.arb-silva.de/>]); and 3) relevant project data that include i) an alignment of sequence data from cultured endophytes generated by an NSF Dimensions of Biodiversity project (<http://www.endobiodiversity.org/>) (Arnold et al., 2009; U'Ren et al., 2009; U'Ren et al., 2010; U'Ren, 2011; U'Ren et al., 2012), ii) sequences of endophytes isolated in culture in our study, and fruiting bodies collected in this study, and iii) amplicon sequences generated by our study. To make sure data from public databases only contain sequences of interest, sets of primers were used to filter unwanted sequences using cutadapt (Martin, 2011) with error allowance of 10%. For sequences from GenBank, only sequences containing both primers ITS4 (White et al., 1990) and LR3 or LR0R and LR15 (Vilgalys unpublished [<http://www.botany.duke.edu/fungi/mycolab>]) were kept to reduce the chance of random matching. Two reference datasets were used to decrease the possibility of deleting correct sequences because one of the primers was trimmed out. Only one primer set (LR0R+LR3) was used to check sequences of the Silva database because we expected this database to contain fewer contaminations. For the LR0R and LR3 libraries, the LR0R and LR3 primers were trimmed, respectively. After the primers were trimmed, the first 250 bp were kept for the LR0R dataset and the last 250 bp were kept for the LR3 dataset. This step ensured that sequences in the reference libraries had similar length, and the length of 250 bp was used

because the RNA-seq reads were 125 bp pair-ended. Our LR0R database included 69,999 reference sequences (hereafter referred to as our LR0R primary database) and our LR3 database included 48,930 sequences (hereafter referred to as our LR3 primary database; see Supplementary Fig S2a, S3a).

After quality filtering, the remaining reads (368,703,726 reads) of the metatranscriptome were mapped against the library described above. All reads mapped to the primary library with Bowtie2 (Langmead and Salzberg, 2012) were pooled and used to create the secondary reference databases. Two methods were applied for the construction of the secondary reference databases. The first method used Velvet (Zerbino and Birney, 2008) to assemble the mapped reads. The second method used USEARCH (Edgar, 2013) to merge paired-end reads. Hereafter, databases created through the Velvet assemblage process will be referred to as MA (Metatranscriptome Assembled) datasets, and those created through USEARCH pair-merging will be referred to as MP (Metatranscriptome Pair-merged) datasets.

Alignments of the representative sequences were manually checked with Mesquite (Maddison and Maddison, 2011) to confirm that they represented LR0R or LR3 regions. These sequences also were BLASTed against the NCBI database to confirm that the sequences represented fungi. All unreliable sequences or regions were removed.

Methods S7. Comparison between metatranscriptome and amplicon sequencing data for the study of fungal communities associated with plants

Amplicon sequencing had deeper and more even sampling depths across individual samples according to rarefaction curves (Supplementary Fig S10).

Metatranscriptomic (shotgun sequencing) and amplicon-based approaches each have their own advantages and disadvantages (Lindahl et al., 2013).

Metatranscriptomes provide researchers flexibility to test hypotheses with many molecular markers. Not only the presence/absence of OTUs is detected, but rRNA read abundances can be used as proxies for levels of metabolic activity for OTUs. Moreover, after identifying fungi of interest using rRNA and their levels of metabolic activity, their mRNA can be studied to explore their function. A shotgun based

metatranscriptomic approach also avoids primer and PCR biases (Lindahl et al., 2013). However, detection of rRNA through mRNA-enriched metatranscriptomics is a relatively low throughput method for high complexity communities as their rarefaction curves showed (Supplementary Fig S10). Construction of comprehensive but clean reference databases requires work and experience, but nevertheless can lead to reference dataset biases. Furthermore, metatranscriptomic data are usually obtained exclusively for mRNA-based studies, thus mRNA enrichment steps (e.g., polyA enrichment and rRNA depletion methods) usually are applied for library construction. Yet untested, these steps might introduce biases and are selective for certain taxa. In turn, amplicon sequencing methods are cheaper and have a higher coverage per unit sample, but primer design and PCR steps can introduce biases (Lindahl et al., 2013; Nguyen et al., 2015).

The metatranscriptome pair-merging (MP) method yielded consistent results for primers LR0R and LR3 (Figs 2-5), whereas the metatranscriptome assembled (MA) approach provided results that differed between LR0R and LR3. For example, the ratios of reads mapped (Supplementary Fig S2b, S3b) and the major OTUs detected (Fig 5) by MA differed between LR0R and LR3, suggesting that the MA method is more biased given the selected markers. The MP method is also less likely to create chimeric sequences in the secondary database, because two pairs should always come from the same fragment during the sequencing process. Currently, the reference sequences generated with pair-merging (MP) strategies are shorter (median length of LR0R-MA = 176 bp vs. LR0R-MP = 123 bp; LR3-MA = 195 bp vs. LR3-MP = 136 bp). With next-generation sequencing technology improving at a fast pace we expect this difference in length to become irrelevant in the near future (Lindahl et al., 2013). In terms of marker choice (LR0R vs. LR3), we note that some taxa recovered by one dataset were not always recovered by others; thus, having several datasets can lower the chances of misinterpreting the data due to reference data bias and/or stochasticity. Also, different markers for the same fungus might assign it to different taxonomic ranks, which is very helpful when establishing the taxonomic composition of fungal communities.

One problem we encountered was the difficulty to assign informative taxonomy to each OTU. Because sequences were relatively short (~120-200 bp), robust assignments often were not possible at the genus level (e.g. ~40% OTUs for LR0R_MP dataset were only assigned to hierarchies higher than the genus level), making biological comparisons difficult. An additional concern is the inadequate database coverage and taxonomic information for many fungal OTUs. Even in the case where we have longer sequences available for culturable fungi, the majority of their BLAST top hits only have taxonomic information at high ranks (e.g., class), impeding further taxonomic comparisons. We used Megablast + MEGAN (Huson et al., 2007) to report taxonomic assignments, coupled with the RDP classifier utilizing the SILVA database (Liu et al., 2012) (Supplementary Tables S7-S12). Overall the results from these two taxonomy-assignment approaches were similar. However, for some sample sets, one package would provide a more detailed taxonomic assignment than the other. For example, some taxa assigned to *Rickenella* by RDP were only assigned to Agaricomycetes/Tricholomataceae by Megablast + MEGAN (Supplementary Tables S7-S8, S10 and S12). The OTUs classified as *Cortinarius* (Figs 3 and 5) by MEGAN may be *Galerina*, because 1) the BLAST top hit is *Galerina*, and 2) *Galerina* fruiting bodies were commonly collected at the same field site. When considering the OTU delimitation threshold, the results based on OTUs using a 95% similarity threshold (Supplementary Fig S9) detected similar results compared to those based on a 97% similarity threshold (Figs 2-6). While the ITS (internal transcribed spacer) region (Schoch et al., 2012) is likely to provide a better taxonomic assignment, there are far fewer reads from the ITS region in metatranscriptomic datasets than from the LSU region (Liao et al., 2014). Because ITS transcripts are believed to be degraded quickly (Liang and Fournier, 1997), the ITS region may not be an ideal marker for studying fungal (ribosomal) activity when using metatranscriptomic data.

Differentially Expressed Gene detection (Love et al., 2014) worked well to detect differentially active taxa for our dataset. A similar approach was recently implemented in Qiime (Caporaso et al., 2010), but as pointed out in their

documentation (http://qiime.org/scripts/differential_abundance.html), this application has not been tested broadly and merits caution.

A previous metatranscriptomic analysis of ectomycorrhizae revealed dominance by only a few fungi and suggested that fungal communities are simpler than those suggested by rDNA amplicon sequencing (Liao et al., 2014). Our metatranscriptomic analyses showed that the active fungal communities in association with *D. scoparium* are complex, that metatranscriptomic methods are applicable to the study of complex communities in environmental samples, and that the main taxa detected through shotgun sequencing (with fewer PCR amplifications) are similar to those revealed by cDNA based amplicon sequencing. A purely metatranscriptomic approach detected a higher proportion of active fungi represented by certain taxa (e.g., Mortierellomycotina in Fig 4) compared to the amplicon-based method, possibly due to PCR efficiency and primer bias. This suggests that the contributions of some taxa might be underestimated when using only an amplicon-sequencing method (Teder et al., 2015), a question worthy of further study. Notably, metatranscriptomic data generally have lower sequencing depth (Supplementary Fig S10). Further discussion of methodological approaches is provided in Supplementary Method S7 and forming the basis for future work.

References for Supplementary Methods

- Arnold AE, Miadlikowska J, Higgins KL, Sarvate SD, Gugger P, Way A, Hofstetter V, Kauff F, Lutzoni F. 2009.** A phylogenetic estimation of trophic transition networks for ascomycetous fungi: Are lichens cradles of symbiotrophic fungal diversification? *Syst Biol* **58**(3): 283-297.
- Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A, the Galaxy T. 2010.** Manipulation of FASTQ data with Galaxy. *Bioinformatics* **26**(14): 1783-1785.
- Bolger AM, Lohse M, Usadel B. 2014.** Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**(15): 2114-2120.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JL, et al. 2010.** QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**(5): 335-336.
- Edgar RC. 2013.** UPPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* **10**(10): 996-998.
- Hinrikson HP, Hurst SF, Lott TJ, Warnock DW, Morrison CJ. 2005.** Assessment of ribosomal large-subunit D1-D2, internal transcribed spacer 1, and internal transcribed spacer 2 regions as targets for molecular identification of medically important aspergillus species. *Journal of Clinical Microbiology* **43**(5): 2092-2103.
- Huson DH, Auch AF, Qi J, Schuster SC. 2007.** MEGAN analysis of metagenomic data. *Genome Research* **17**(3): 377-386.
- Langmead B, Salzberg SL. 2012.** Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**(4): 357-359.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup G. 2009.** The sequence alignment/map format and SAMtools. *Bioinformatics* **25**(16): 2078-2079.
- Liang W-Q, Fournier MJ. 1997.** Synthesis of functional eukaryotic ribosomal RNAs in trans: Development of a novel *in vivo* rDNA system for dissecting ribosomal biogenesis. *Proceedings of the National Academy of Sciences* **94**(7): 2864-2868.
- Liao H-L, Chen Y, Bruns TD, Peay KG, Taylor JW, Branco S, Talbot JM, Vilgalys R. 2014.** Metatranscriptomic analysis of ectomycorrhizal roots reveal genes associated with *Piloderma-Pinus* symbiosis: Improved methodologies for assessing gene expression *in situ*. *Environmental Microbiology* **16**(12): 3730-3742.
- Lindahl BD, Nilsson RH, Tedersoo L, Abarenkov K, Carlsen T, Kjølner R, Kõljalg U, Pennanen T, Rosendahl S, Stenlid J, et al. 2013.** Fungal community analysis by high-throughput sequencing of amplified markers – a user's guide. *New Phytologist* **199**(1): 288-299.
- Liu K-L, Porras-Alfaro A, Kuske CR, Eichorst SA, Xie G. 2012.** Accurate, rapid taxonomic classification of fungal large-subunit rRNA genes. *Applied and Environmental Microbiology* **78**(5): 1523-1533.

- Love M, Huber W, Anders S. 2014.** Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**(12): 550.
- Maddison WP, Maddison DR. 2011.** Mesquite: a modular system for evolutionary analysis. Version 2.75 <http://mesquiteproject.org/>.
- Martin M. 2011.** Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*; Vol 17, No 1: *Next Generation Sequencing Data Analysis*.
- Nguyen NH, Smith D, Peay K, Kennedy P. 2015.** Parsing ecological signal from noise in next generation amplicon sequencing. *New Phytologist* **205**(4): 1389-1393.
- Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Fungal Barcoding C. 2012.** Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences* **109**(16): 6241-6246.
- Tedersoo L, Anslan S, Bahram M, Põlme S, Riit T, Liiv I, Kõljalg U, Kisand V, Nilsson H, Hildebrand F, et al. 2015.** Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *Mycologia* **10**: 1-43.
- U'Ren JM, Dalling JW, Gallery RE, Maddison DR, Davis EC, Gibson CM, Arnold AE. 2009.** Diversity and evolutionary origins of fungi associated with seeds of a neotropical pioneer tree: a case study for analysing fungal environmental samples. *Mycological Research* **113**(4): 432-449.
- U'Ren JM, Lutzoni F, Miadlikowska J, Arnold AE. 2010.** Community analysis reveals close affinities between endophytic and endolichenic fungi in mosses and lichens. *Microbial Ecology* **60**(2): 340-353.
- U'Ren JM. 2011.** *Host-, geographic-, and ecological specificity of endophytic and endolichenic fungal communities*. PhD thesis. University of Arizona USA.
- U'Ren JM, Lutzoni F, Miadlikowska J, Laetsch AD, Arnold AE. 2012.** Host and geographic structure of endophytic and endolichenic fungi at a continental scale. *American Journal of Botany* **99**(5): 898-914.
- Vilgalys R, Hester M. 1990.** Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several *Cryptococcus* species. *Journal of Bacteriology* **172**(8): 4238-4246.
- White TJ, Bruns T, Lee S, Taylor J 1990.** 38 - Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ eds. *PCR Protocols*. San Diego: Academic Press, 315-322.
- Zerbino DR, Birney E. 2008.** Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research* **18**(5): 821-829.

Supporting Information Figures S1-S10.

Fig. S1. Quality filtering of cDNA amplicon sequencing.

Amplicon sequencing results for LR0R (forward) and LR3 (reverse) reads and read numbers after the removal of adaptors and quality filtering. The Y-axis corresponds to the read number. The X-axis was labeled according to the layers of *D. scoparium* and sampling replicates (Fig 1).

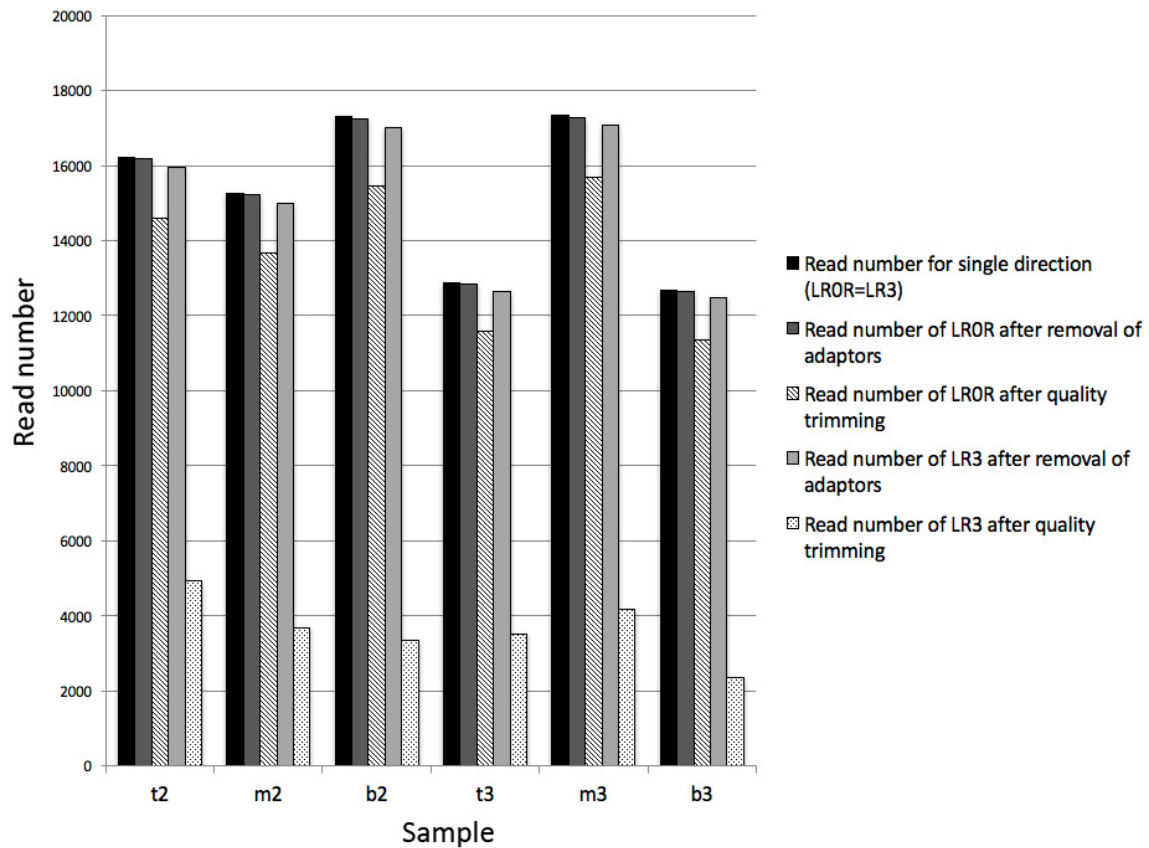
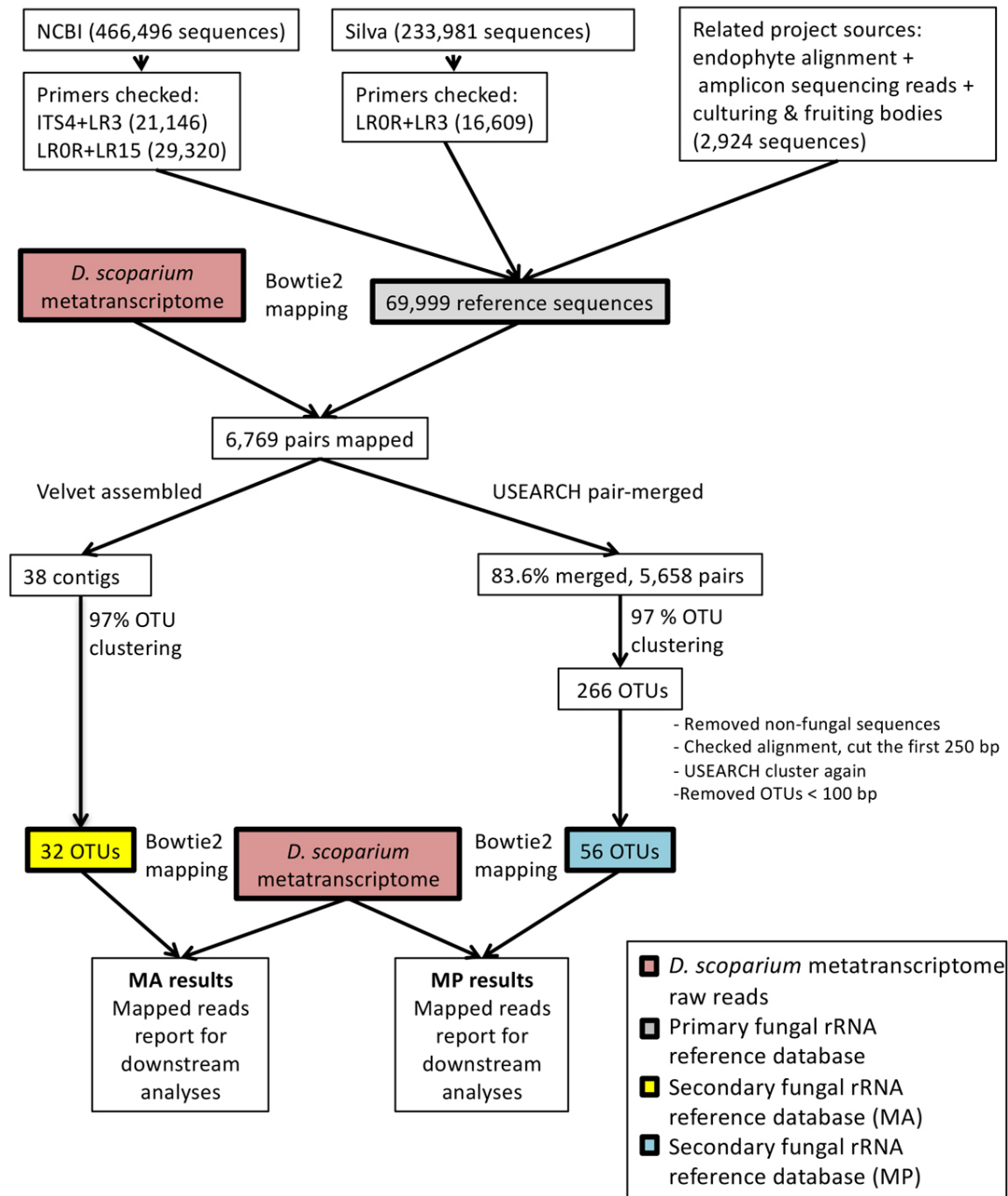


Fig. S2. Workflow for metatranscriptome LR0R database.

(a) Workflow for LR0R database constructions. The primary fungal database refers to rRNA reference sequences collected from NCBI GenBank, Silva database and project related resources. Secondary fungal rRNA reference databases were created using the reads mapped to the primary database from the metatranscriptome. (b) Histogram showing ratios of fungal reads mapped to the three LR0R libraries. The libraries were color-coded according to Figure S2a. The Y-axis corresponds to the ratio of reads mapped to the three LR0R libraries. The X-axis is labeled according to the layers of *D. scoparium* and sampling replicates (Fig. 1).

(a) LROR database (≈250 bp, sequences <100 bp removed)



(b)

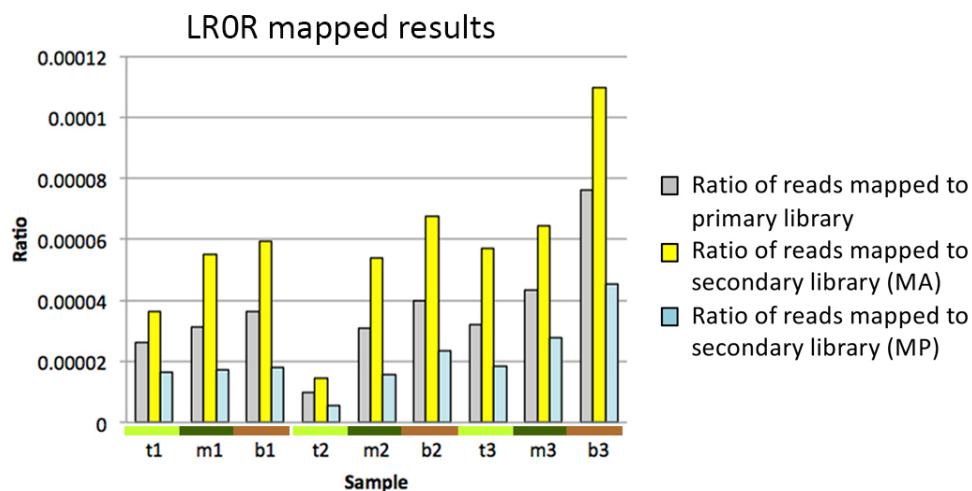


Fig. S3. Workflow for metatranscriptome LR3 database.

(a) Workflow for LR3 database constructions. The primary fungal database refers to rRNA reference sequences collected from NCBI GenBank, Silva database and project related resources. Secondary fungal rRNA reference databases were created using the reads mapped to the primary database from metatranscriptome. (b) Histogram showing ratios of fungal reads mapped to the three LR3 libraries. The libraries were color coded according to Supplementary Fig S2a. The Y-axis corresponds to the ratio of reads mapped to the three LR3 libraries. The X-axis is labeled according to the layers of *D. scoparium* and sampling replicates (Fig 1).

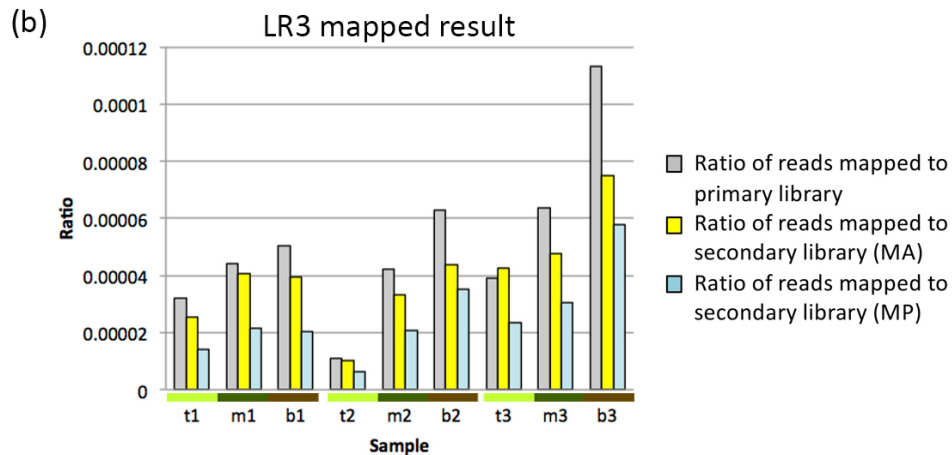
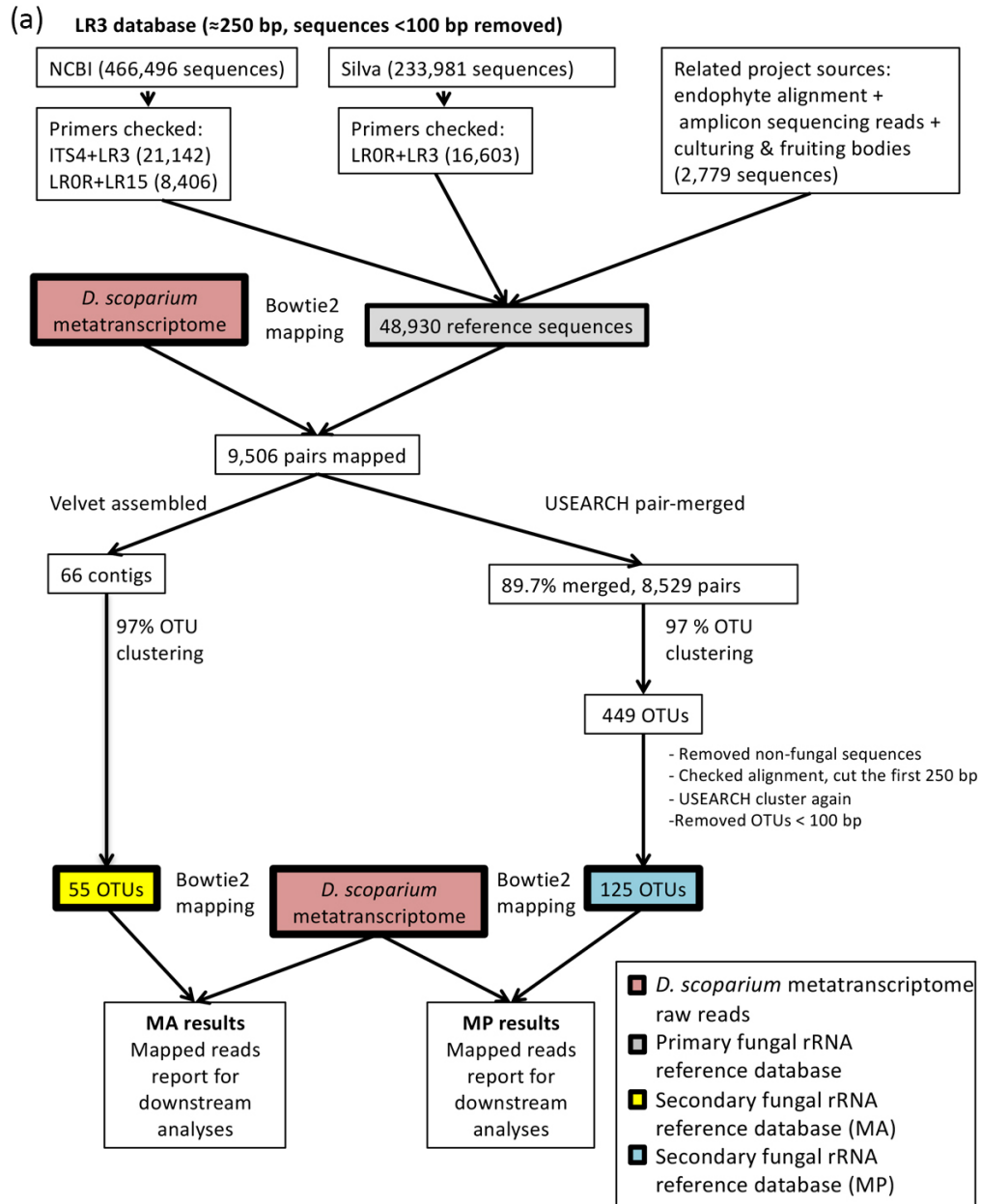


Fig. S4. Ratio of the metatranscriptomic data mapping to the reference plant genome/transcriptome.

Colors correspond to the sampled layer as shown in Figure 1a. The letters t, m and b stand for top, middle and bottom layers. Numbers 1, 2 and 3 reflect replicates shown in Figure 1b. The ratios of reads mapped to the plant *D. scoparium* always declined toward the bottom layer.

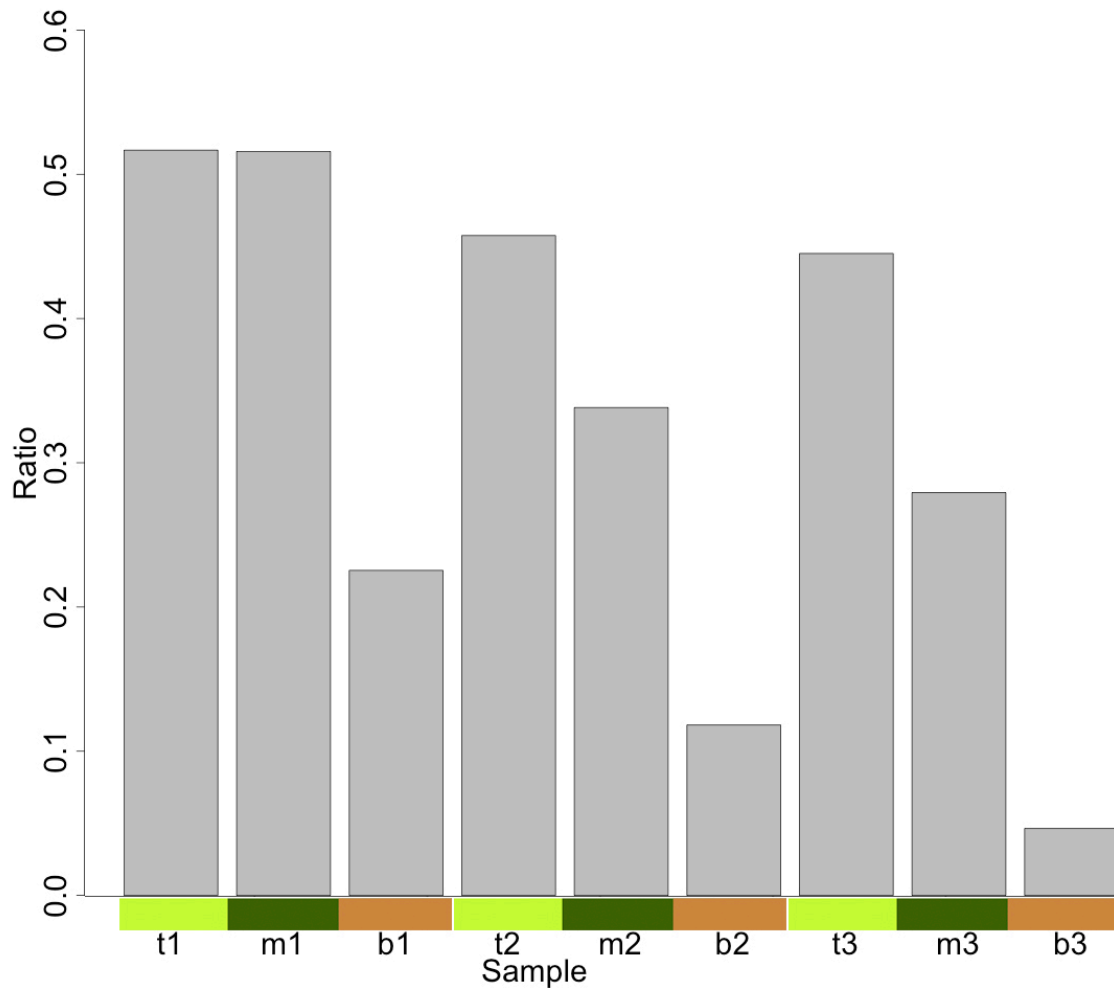


Fig. S5. Isolation frequency of fungal and bacterial endophytes in culture.

Colors correspond to the sample layer as shown in Fig 1a. The letters t, m and b stand for top, middle and bottom layers. Numbers 1, 2 and 3 reflect microsites shown in Fig 1b. Many bacteria were isolated, especially in the bottom layer of the gametophytes. These bacterial communities will be studied in a separate project.

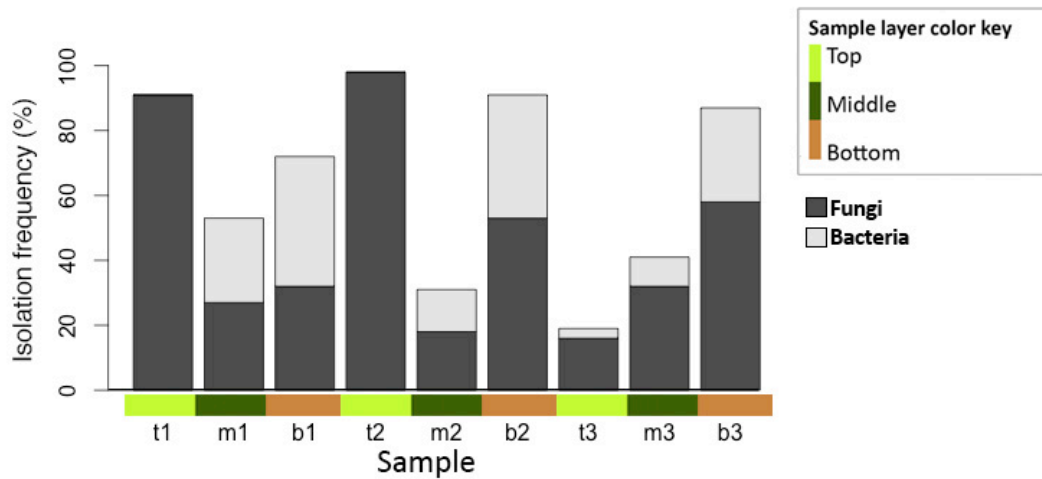
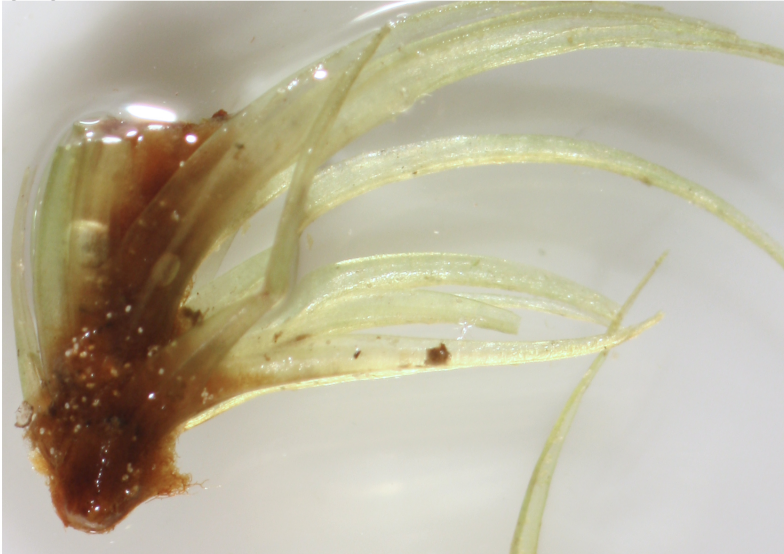


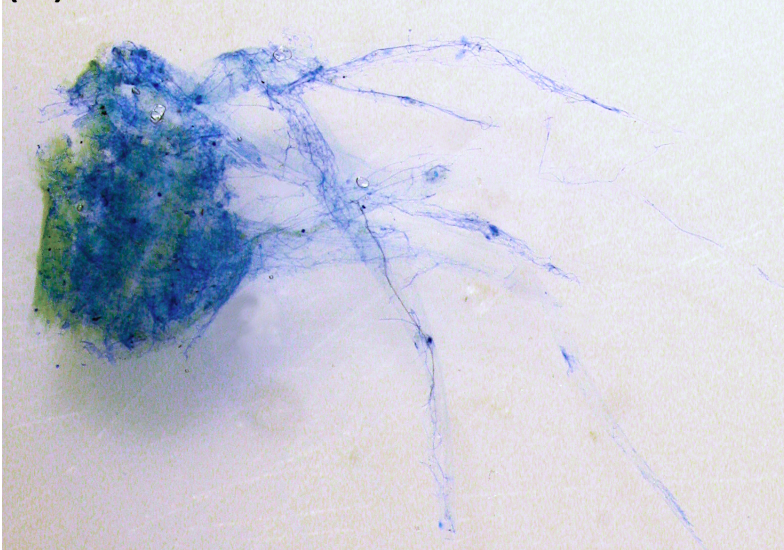
Fig. S6. Fungal hyphae in photosynthetic plant tissues.

The moss was washed vigorously with tap water to remove epiphytic debris prior to microscopic examinations. (a) Photosynthetic tissue of *D. scoparium* under dissecting microscope before KOH and staining procedure. (b) Photosynthetic tissue of *D. scoparium* under dissecting microscope after being cleared with 3% KOH overnight and stained with Trypan blue.

(a)



(b)



1 mm

Fig. S7. Fruit-bodies collected at the sampling site.

Fruit-bodies collected at sampling site (Fig 1b) that were attached to gametophytes of *Dicranum scoparium*.



Fig. S8. Results from the LR3-MA dataset when using an OTU delimitation of 95% similarity (MA = Metatranscriptome Assembled).

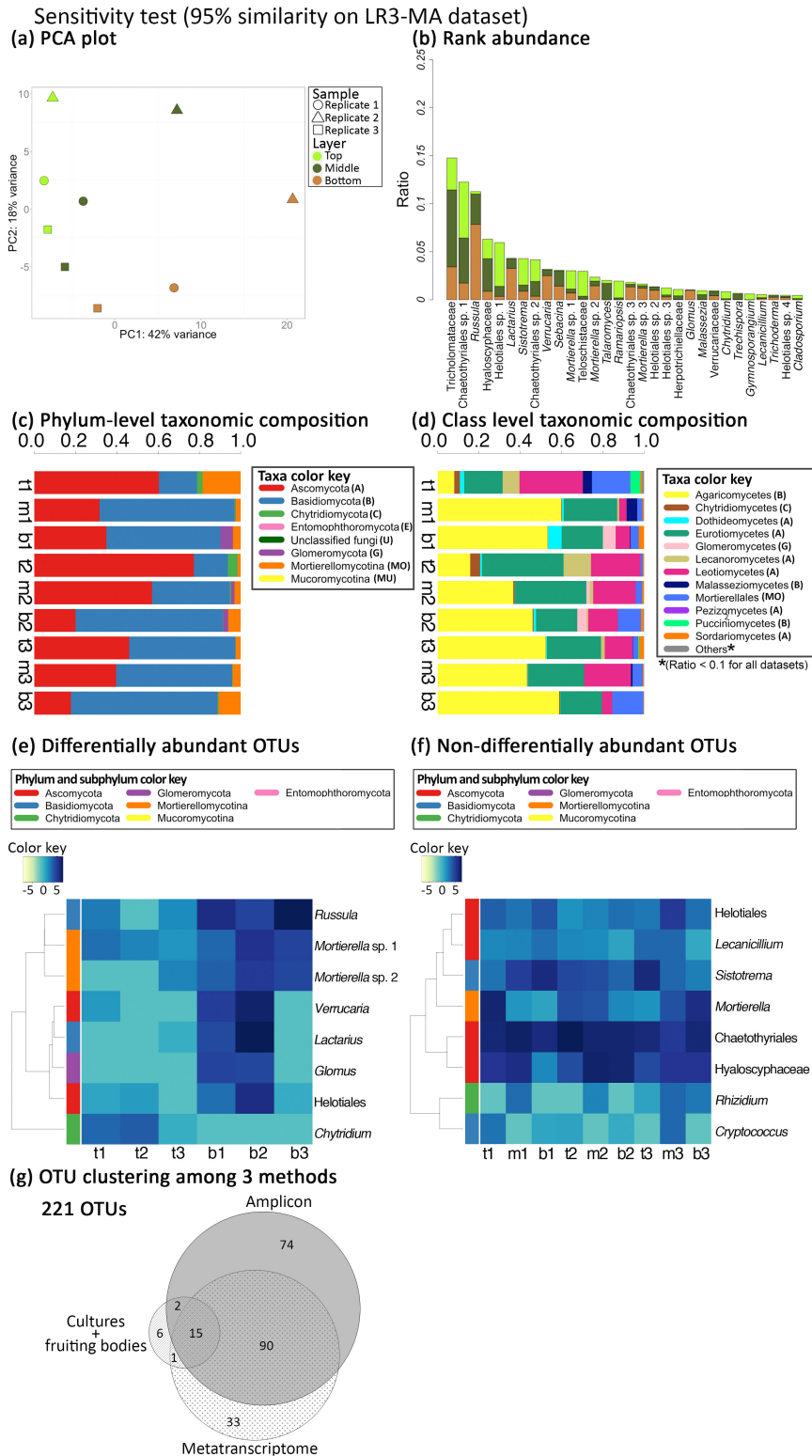


Fig. S9. Ratio of reads mapped to Ascomycota, Basidiomycota, and non-Dikarya phyla based on the LR0R-MP dataset (MP= Metatranscriptome Pair-merged).

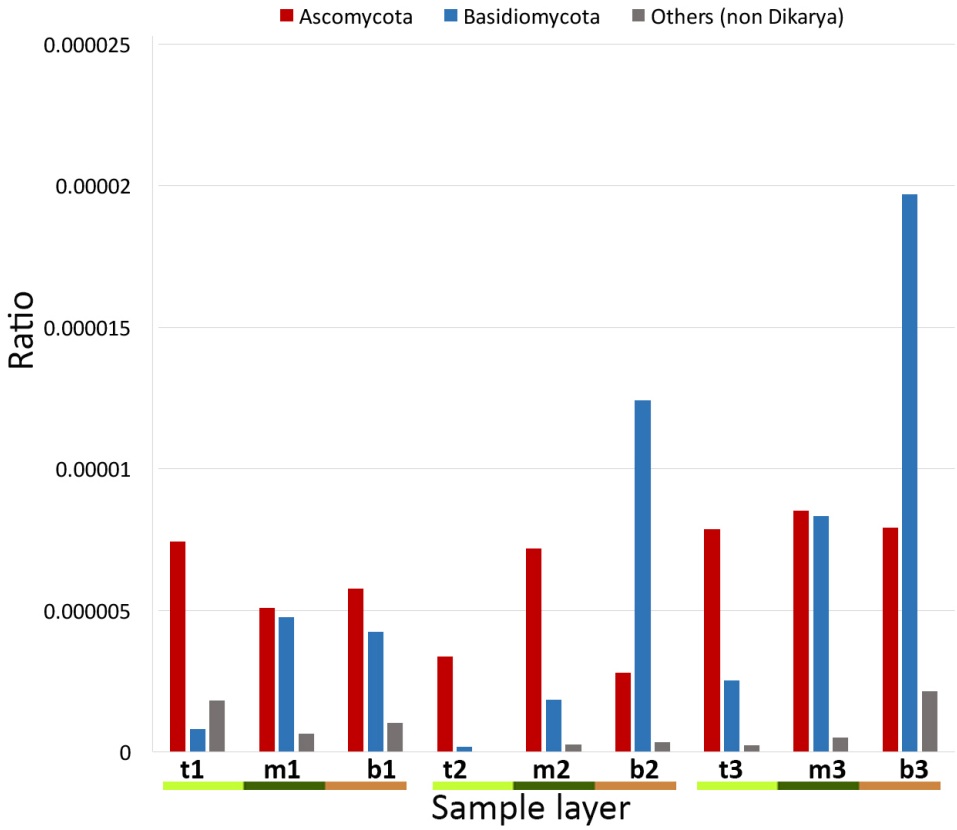


Fig. S10. Rarefaction curves of all datasets. (MA = Metatranscriptome Assembled, MP= Metatranscriptome Pair-merged, amp = amplicon sequencing)

