



Comparative Fungal Community Analyses Using Metatranscriptomics and Internal Transcribed Spacer Amplicon Sequencing from Norway Spruce

Andreas N. Schneider,^a John Sundh,^b Görel Sundström,^c Kerstin Richau,^a Nicolas Delhomme,^c Manfred Grabherr,^d Vaughan Hurry,^c Nathaniel R. Street^a

^aUmeå Plant Science Centre, Department of Plant Physiology, Umeå University, Umeå, Sweden

^bDepartment of Biochemistry and Biophysics, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Stockholm University, Solna, Sweden

^cUmeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish Agricultural University, Umeå, Sweden

^dDepartment of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden

ABSTRACT The health, growth, and fitness of boreal forest trees are impacted and improved by their associated microbiomes. Microbial gene expression and functional activity can be assayed with RNA sequencing (RNA-Seq) data from host samples. In contrast, phylogenetic marker gene amplicon sequencing data are used to assess taxonomic composition and community structure of the microbiome. Few studies have considered how much of this structural and taxonomic information is included in transcriptomic data from matched samples. Here, we described fungal communities using both host-derived RNA-Seq and fungal ITS1 DNA amplicon sequencing to compare the outcomes between the methods. We used a panel of root and needle samples from the coniferous tree species *Picea abies* (Norway spruce) growing in untreated (nutrient-deficient) and nutrient-enriched plots at the Flakaliden forest research site in boreal northern Sweden. We show that the relationship between samples and alpha and beta diversity indicated by the fungal transcriptome is in agreement with that generated by the ITS data, while also identifying a lack of taxonomic overlap due to limitations imposed by current database coverage. Furthermore, we demonstrate how metatranscriptomics data additionally provide biologically informative functional insights. At the community level, there were changes in starch and sucrose metabolism, biosynthesis of amino acids, and pentose and glucuronate interconversions, while processing of organic macromolecules, including aromatic and heterocyclic compounds, was enriched in transcripts assigned to the genus *Cortinarius*.


IMPORTANCE A deeper understanding of microbial communities associated with plants is revealing their importance for plant health and productivity. RNA extracted from plant field samples represents the host and other organisms present. Typically, gene expression studies focus on the plant component or, in a limited number of studies, expression in one or more associated organisms. However, metatranscriptomic data are rarely used for taxonomic profiling, which is currently performed using amplicon approaches. We created an assembly-based, reproducible, and hardware-agnostic workflow to taxonomically and functionally annotate fungal RNA-Seq data obtained from Norway spruce roots, which we compared to matching ITS amplicon sequencing data. While we identified some limitations and caveats, we show that functional, taxonomic, and compositional insights can all be obtained from RNA-Seq data. These findings highlight the potential of metatranscriptomics to advance our understanding of interaction, response, and effect between host plants and their associated microbial communities.

Citation Schneider AN, Sundh J, Sundström G, Richau K, Delhomme N, Grabherr M, Hurry V, Street NR. 2021. Comparative fungal community analyses using metatranscriptomics and internal transcribed spacer amplicon sequencing from Norway spruce. *mSystems* 6:e00884-20. <https://doi.org/10.1128/mSystems.00884-20>.

Editor Ryan McClure, Pacific Northwest National Laboratory

Copyright © 2021 Schneider et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Nathaniel R. Street, nathaniel.street@umu.se.

 Assembly-based analysis of fungal metatranscriptomic data allows for structural, functional, and taxonomic community analyses, of which only the latter is limited by current availability of fungal reference genomes.

Received 1 September 2020

Accepted 27 January 2021

Published 16 February 2021

KEYWORDS fungi, metatranscriptomics, ITS amplicon sequencing, Norway spruce, nutrient enrichment, ectomycorrhiza, tree roots, phyllosphere, fungi, phyllosphere-inhabiting microbes

A growing body of research shows that plants harbor a complex assemblage of epiphytic and endophytic symbionts (1). Understanding the composition and role of the microbial components of such systems raises fundamental questions concerning the taxonomic composition and biological functions provided by these communities and how they influence plant survival and fitness. High-throughput DNA sequencing technologies have vastly improved our ability to assay these complex and diverse microbial communities (2, 3). The current *de facto* standard of metagenomics is the use of amplicons spanning regions of marker genes, usually the internal transcribed spacer (ITS) region of the rRNA gene for fungal species and variable regions of the 16S rRNA gene for bacteria (4, 5), for both of which extensive reference databases exist (6, 7). Although useful for taxonomic profiling, DNA-based amplicon methods suffer from methodological biases such as not accounting for multiple rRNA copies per cell and preferential primer binding, leading to bias for or against certain taxa (8, 9). Moreover, DNA-based methods cannot differentiate between living and dead sources of DNA (10). In contrast to examining DNA, RNA sequencing (RNA-Seq) captures actively expressed sequences as well as their relative abundance. Among the numerous appealing qualities of RNA-Seq is the nearly universal coverage of the transcriptome. In the current context, and particularly in the case of field samples, that coverage comprises transcripts from both a host and its associated microbial community, enabling what we previously referred to as serendipitous metatranscriptomics (11) and what others have termed tripartite sequencing (12) or, in the case of a host and single microbial species, dual RNA-Seq (13). One characteristic of using metatranscriptome (i.e., the transcriptome of a whole community) data is that it yields insight into the biological processes active within microbial communities, providing functional insights (14–18). The availability of functional information from both components of the holobiont system (the assemblage of the plant host and the hosted microbial community) could be transformative in advancing our understanding of the development, dynamics, interactions, and effects of these two components (17, 19, 20). In principle, RNA-Seq applied to a holobiont system would enable taxonomic profiling of the represented species, offering taxonomic information in addition to information on the biological processes actively represented in the metatranscriptome.

There have been few systematic comparisons of metatranscriptomics data to those from amplicon sequencing of the 16S or ITS regions of rRNA genes. One study using human stool samples concluded that total metatranscriptome data have higher sensitivity and reproducibility than both ITS and 16S amplicon data (21). More such studies are needed to understand whether both methods provide similar insight into community diversity, species composition, and biological function. Here, this question was addressed by performing a comparison of taxonomic information and community structure obtained from mRNA-based metatranscriptomics and amplicon sequencing of the fungal ITS1 region. As a study system, we used a panel of root and needle samples from *Picea abies* (Norway spruce) growing in northern boreal Sweden. The boreal forest covers around one-third of the world's forested areas and is mostly characterized by harsh climates and N-limited plant growth (22). These forests are dominated by conifers, and they host complex communities of microorganisms, both in the soil and in close association with the forest trees. Ectomycorrhizal (ECM) fungi are especially important in this context, colonizing over 90% of root tips in the boreal forest (23). ECM enhance tree nutrient uptake and are important drivers of carbon and nutrient cycling in the boreal forests (24). Another ubiquitous and important group of fungi are saprotrophs, which play an important role in degrading organic litter and root detritus (25). The site sampled in this study is part of a controlled short-term (5 years) and long-term (25 years) nutrient enrichment (NE) experiment including untreated nutrient-deficient

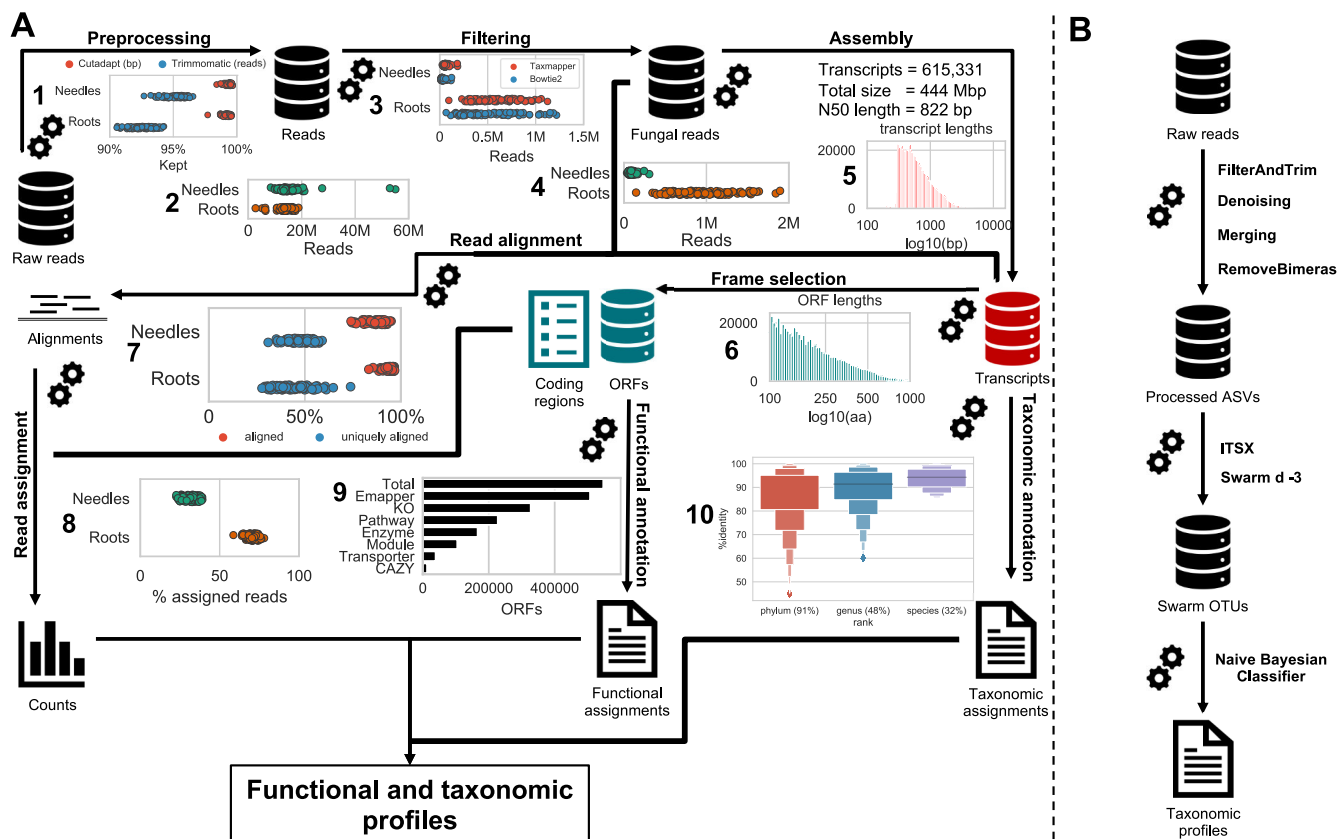


FIG 1 (A) Overview of the RNA-Seq workflow. (Step 1) The inset shows the proportion of the raw data that was kept after trimming adapters and removing low-quality regions (shown in base pairs for cutadapt and as reads for Trimmomatic). (Step 2) Reads remaining after preprocessing in needle and root samples. (Step 3) The inset shows the number of reads identified as fungal by bowtie2 alignments and TaxMapper assignments in needle and root samples. (Step 4) Number of fungal reads in needle and root samples after filtering. (Step 5) The inset shows length distribution of assembled transcripts (log₁₀ scale) after assembly using Megahit. (Step 6) Length distribution of the open reading frames (ORFs) as determined by GeneMarkS-T (in amino acids, log₁₀ scale). (Step 7) The inset shows overall and uniquely aligned fraction of reads for needle and root samples, after aligning fungal reads to the assembled transcripts using bowtie2. (Step 8) Fraction of assigned reads in needle and root samples, as determined by FeatureCounts. (Step 9) The inset shows number of total ORFs and number of ORFs with different levels of functional annotations obtained from eggno-mapper. (Step 10) Distribution of amino acid identity for DIAMOND BLASTX hits used to assign taxonomy at the different ranks, using the contigtax tool. The fraction of transcripts assigned at each rank is shown in parentheses on the x axis. Full statistics on surviving reads are available in Table S1. (B) Overview of the amplicon sequencing workflow. Raw, demultiplexed reads were filtered and trimmed, after which dada2 was used to denoise the reads and to merge forward and reverse reads. Subsequently, chimeras were removed, and the resulting amplicon sequencing variants (ASVs) were cut with ITSx and clustered into Swarm operational taxonomic units (SOTUs). Finally, taxonomy was assigned to SOTUs using the dada2 naive Bayesian classifier and the UNITE database. Detailed read counts can be found in Table S1.

(ND) control plots (26, 27). The aim of this study was 3-fold: (i) to implement a bioinformatic workflow for metatranscriptomic RNA-Seq data that filters host-derived reads and assigns taxonomic and functional annotations to assembled fungal transcripts; (ii) to compare results derived using this pipeline to rRNA gene amplicon-based data of the fungal ITS region (28); and (iii) to demonstrate the potential of our metatranscriptomic data for providing multifaceted, functional insight into actively expressed genes, for example, identifying biological processes that are enriched in response to long-term NE both at community level and for the selected genus *Cortinarius*.

RESULTS

Pipeline development and data set statistics. RNA-Seq of roots and needles yielded an average 14.7 million reads per sample after adapter/quality trimming (Fig. 1A), of which 0.6% (89,763 reads) and 6.7% (933,229 reads) on average were identified as fungal (by alignment to the JGI MycoCosm and TaxMapper databases) in the needle and root samples, respectively (see Table S1 in the supplemental material). Assembly of fungal reads using Megahit generated 615,331 transcripts, with a total size of 444 Mbp. The length of transcripts ranged from 200 to 12,588, with an N₅₀

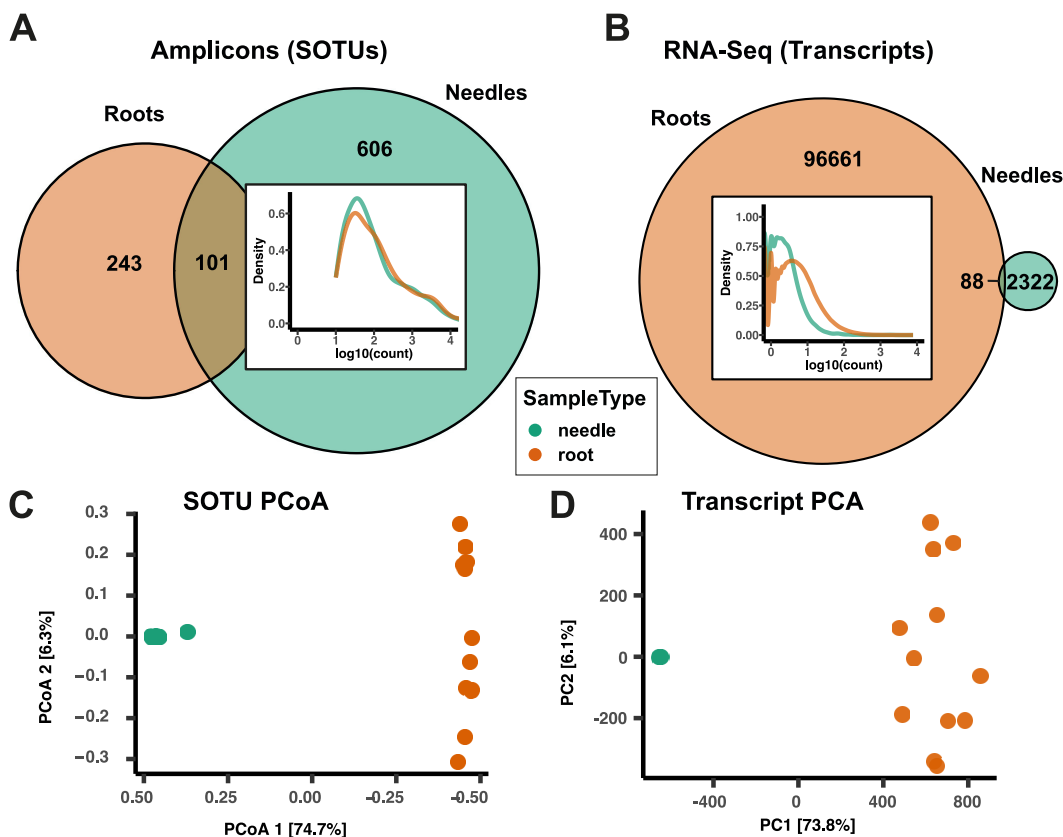


FIG 2 Nutrient deficient (ND) sample overview, contrasting fungal communities in Norway spruce needles and roots. (A and B) Venn diagrams showing the number of postfiltering fungal Swarm operational taxonomic units (SOTUs) (A) and fungal transcripts (B) obtained from root and needle ND samples. Inside the Venn diagrams are density curves showing the \log_{10} -transformed total count distribution in needle (green) and root (brown) samples. (C) Principal-coordinate analysis (PCoA) of Bray-Curtis dissimilarities between needle and root ND samples, obtained from ITS1 amplicon sequencing data. (D) Principal-component analysis (PCA) of variance stabilization-transformed counts from the metatranscriptomes of Norway spruce roots and needles.

length of 822 bp (for an ExN_{50} [i.e., the N_{50} value over the most highly expressed genes that represent $x\%$ of the total normalized expression data] graph and a comparison to Trans-ABYSS and Trinity, see Fig. S1 and Text S1). A total of 547,305 open reading frames (ORFs) were called on the assembled transcripts, with a median length of 98 amino acids, while for 68,029 (11.1%) transcripts no ORF was found. An average of 34.3% reads in needle samples and 70.6% in root samples were aligned to these ORFs. Functional annotation of ORFs was performed, with 92.7% of ORFs having a hit in the eggNOG database, of which 64.5% were assigned to a Kyoto Encyclopedia of Genes and Genomes (KEGG) ortholog and 59.8% were assigned Gene Ontology (GO) terms. Taxonomic assignments resulted in 95.5%, 50.4%, and 34.2% of transcripts assigned at the phylum, genus, and species levels, respectively. A more detailed description is available in Text S1.

For the ITS1 amplicon sequencing data (Fig. 1B), between 86,279 and 338,800 reads per sample remained (176,734 on average), corresponding to a range between 47 and 78% of the raw reads (Table S1). Denoising and chimera removal resulted in 5,726 ASVs in total, of which 2,694 were found in roots and 3,032 in needle samples. After clustering, there were 2,673 Swarm operational taxonomic units (SOTUs), 1,172 in root samples and 1,890 in needle samples.

Comparison of tree tissues in nutrient deficient control samples. Twice as many SOTUs were observed in the needle ND samples as in roots (Fig. 2A), consistent with the published analysis of these data (28). In contrast, the RNA-Seq data set showed around 50 times more remaining transcripts after abundance filtering in the root than

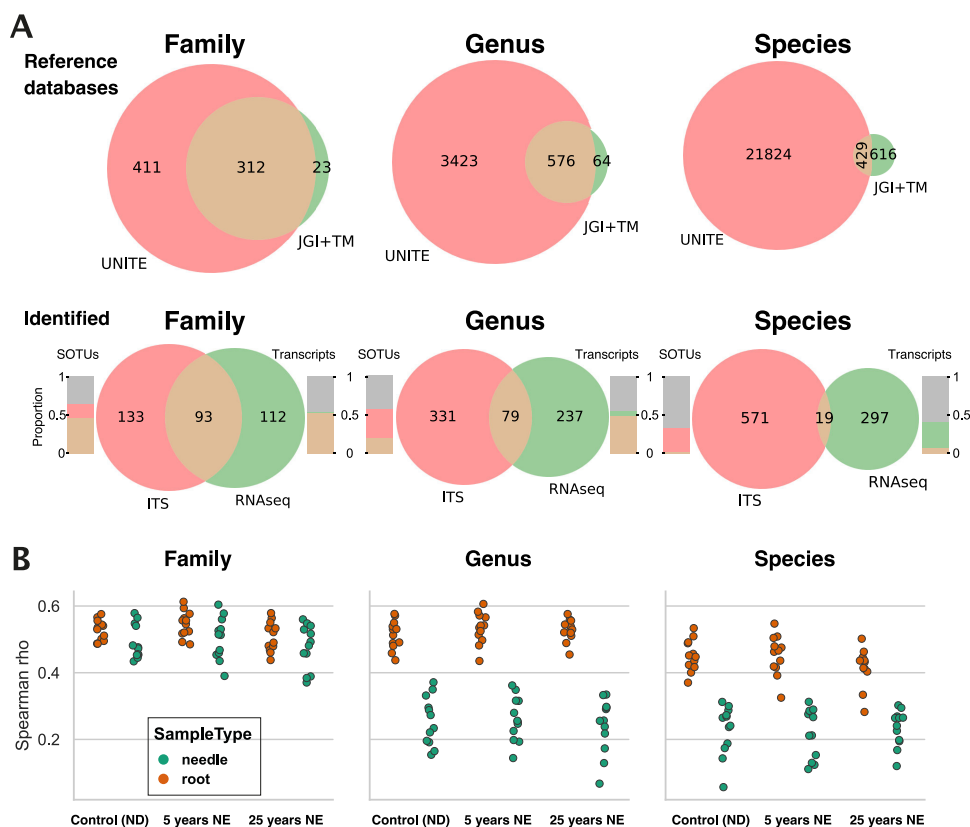


FIG 3 Taxonomic congruence between RNA-Seq and ITS amplicon sequencing data sets. (A) (Top) Venn diagrams showing taxonomic unit overlap between the UNITE database and the JGI MycoCosm and TaxMapper (JGI+TM) genome databases at the family, genus, and species levels. (Bottom) Taxonomic unit overlap of identified taxa in the RNA and the ITS sequencing data sets. Bars indicate the proportion of transcripts and Swarm operational taxonomic units (SOTUs) belonging to the unique and common portions of the Venn diagrams (colors correspond; gray indicates unidentified transcripts/SOTUs on the corresponding taxonomic level). (B) Spearman rank correlations of taxonomic abundance (family, genus, and species levels) between RNA-Seq and ITS amplicon sequencing samples. Colors indicate samples of root (brown) or needle (green) origin.

in the needle samples (Fig. 2B). There was a predominance of low counts per SOTU and transcript in the ITS and the RNA data sets, respectively, particularly in the needle samples (Fig. 2A and B, insets).

A principal-coordinate analysis on the Bray-Curtis dissimilarities between root and needle ND samples in the ITS data set revealed a clear separation of needle and root samples along the first principal coordinate (75% explained variance) (Fig. 2C). The second principal coordinate was characterized by the biggest variation among the root samples, corresponding to the variation between field plots. A principal-component analysis of transcript counts of root and needle ND samples led to a highly similar separation and arrangement of samples (Fig. 2D), but with lower variation among needle samples. The visual congruency between the two data sets for the ND samples was confirmed (Mantel r , 0.89; $P < 0.001$; Procrustes correlation, 0.86; squared m_{12} , 0.26; $P < 0.001$). Due to the small amount of remaining transcripts and the low intersample variance in the needle samples, only root samples were used for later data comparisons.

Comparison of taxonomic annotations in ITS and RNA databases and data sets.

To compare the coverage of the databases used for taxonomic annotation of transcripts (JGI MycoCosm and TaxMapper) and SOTUs (UNITE database), the number of families, genera, and species listed in both or only one of the databases was assessed (Fig. 3A). The proportional overlap between the two databases clearly decreased with lower taxonomic levels. A similar trend was found for taxa identified in the two data sets, but with a lower proportional overlap at the species level than between the

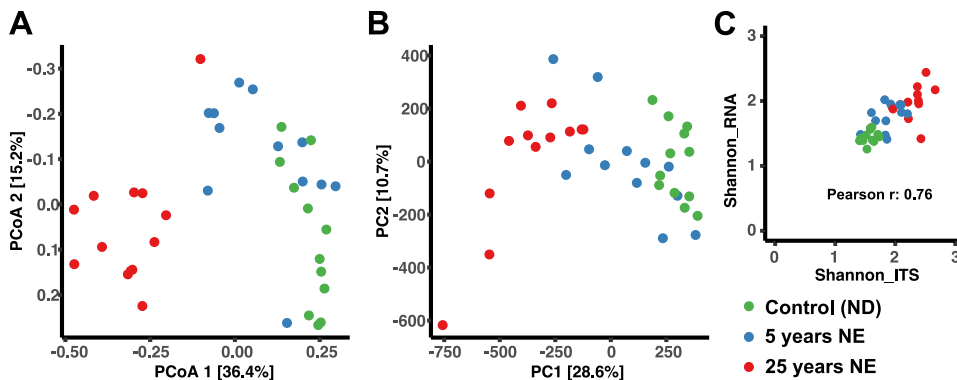


FIG 4 Ordination and alpha diversity index comparison of root samples. (A) Principal-coordinate analysis of rarefied Swarm operational taxonomic unit (SOTU) counts, colored by treatment. (B) Principal-component analysis of variance stabilization transformed transcript counts, colored by treatment. (C) Sample-wise relationship between Shannon diversity index values (genus level) using ITS amplicon sequencing (x axis) and RNA-Seq (y axis), colored by treatment.

databases (Fig. 3A, lower row). At the family level, the percentage of common transcripts and SOTUs was $\sim 50\%$, while at the species level, $<5\%$ were in common (Fig. 3A, area bar graphs). The same trend was apparent on a read count level (Fig. S2A).

To assess how well the relative abundance of common taxa agreed between the RNA and the ITS data sets, Spearman rank correlations were computed for all taxonomic levels (Fig. 3B; Fig. S2B). At the family level, the correlations ranged between 0.4 and 0.6 (medians, 0.53 in roots and 0.48 in needles). In accordance with the Venn diagrams, correlations decreased rapidly at lower taxonomic ranks in needle samples (medians, 0.48 and 0.26 at the genus and species levels, respectively) and moderately in root samples (medians, 0.53 and 0.44 at the genus and species levels, respectively).

Comparison of community structure in root samples. To assess how well the transcript and SOTU data from the root samples structurally correlated with each other, within and between fertilization treatments, ordination analyses were performed. A principal-coordinate analysis (PCoA) on the SOTU counts obtained from the ITS1 sequencing data revealed the same pattern as in the previously published results (Fig. 4A) (28). The 25-year-treated (NE-25) samples clearly separated from the controls on the first principal coordinate (explaining 36% of the variance). A permutational multivariate analysis of variance (PERMANOVA) test showed significance for fertilization treatment ($P < 0.001$) but not for sampling date. Similarly, a principal-component analysis (PCA) of the transcript counts (Fig. 4B) showed that the NE treatment accounted for the highest variance in the data set (29%) and that after 25 years of NE, the fungal transcriptomes in the fertilized plots were distinct from those in the ND samples. PERMANOVA confirmed this ($P < 0.001$), while sampling date was not significant. The correlation between the sample distances in the two ordinations was significant (Mantel r , 0.74; $P < 0.001$; Procrustes correlation, 0.55; squared $m12$, 0.69; $P < 0.001$). Phyllospheric community structure comparisons (Fig. S3) are discussed in the supplemental material (Text S1).

Furthermore, Shannon diversity index values at the genus level between the two data sets were compared (Fig. 4C). The total correlation was strong (Spearman r , 0.76), and the increase in Shannon diversity with longer NE treatment (as reported in reference 28) was apparent in both the ITS and the RNA data (all pairwise comparisons were significant; all $P < 0.02$).

Comparison of highly abundant families and random forest classification by sample source using taxonomic annotations in both data sets. Counts in both data sets were summarized to the family level, and relative proportions of the 12 most abundant family annotations from the two data sets were visualized (Fig. 5A). As expected from the above comparison of taxonomic annotations, the general

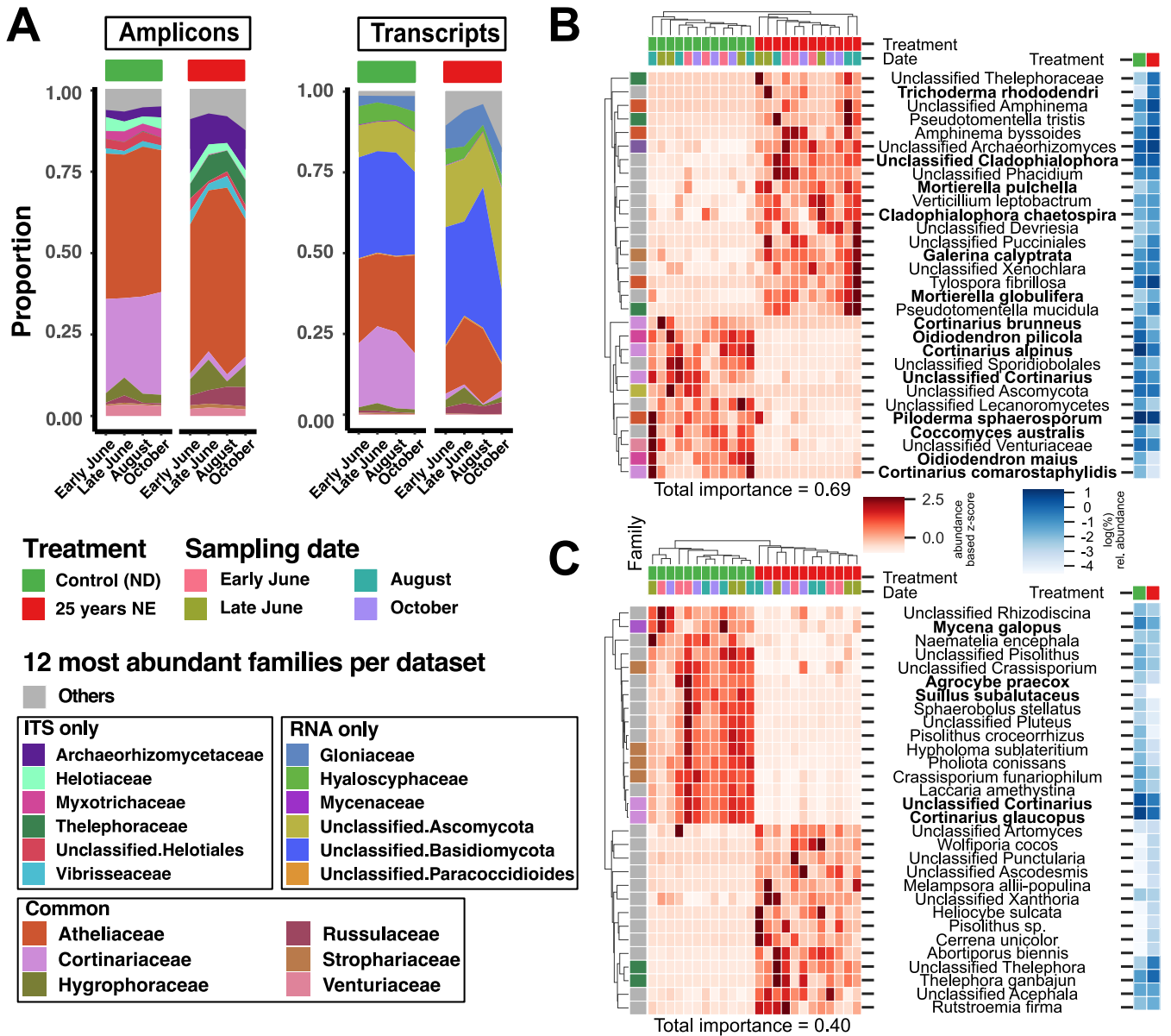


FIG 5 Taxonomic overview and random forest results. (A) Taxonomic overview area plot of ITS amplicon sequencing data (left bars) and annotated transcripts (right bars) from root samples. Counts were summed to the family level, and mean relative abundance per sample type is displayed. Shown in color inside the bars are the 12 most abundant families. Red and green bars on top indicate the treatment; sampling time point is indicated below the columns. (B) ITS data random forest results. Heat map showing the distribution of the 30 most important species (rows) for prediction of samples (columns) into ND/25-year NE groups. Normalized abundance values (summed to the species level) were converted to z-scores per row to highlight differences between samples. Hierarchical clustering of samples and species was performed using correlation metrics and complete linkage clustering. Colors on top indicate treatment and sampling date. Colors in the left margin indicate the corresponding family (gray indicates that the result was not among the 12 most abundant family annotations in any of the two data sets). The two-column heat map to the right indicates average relative abundance in ND and NE samples. Species in bold belong to genera that are found in both data sets. (C) Metatranscriptome random forest results. Heat map showing the distribution of the top 30 most important species (rows) for prediction of samples (columns) into control/25-year-treatment groups. Normalized expression values (summed to species level) were converted to z-scores per row to highlight differences between samples. Hierarchical clustering of samples and species was performed using correlation metrics and complete linkage clustering. Colors on top indicate treatment and sampling date. Colors in the left margin indicate the corresponding family of species. The two-column heat map to the right indicates average relative abundance in control and NE samples. Species in bold belong to genera that are found in both data sets.

overlap was not strong, but notable examples, like Cortinariaceae and Hygrophoraceae, showed very similar abundance distributions. The total proportion occupied by the 12 most abundant family annotations (covering 95% of reads on average) in each data set were highly similar, and this proportion decreased with longer NE treatment.

A random forest classifier on both ITS and RNA data set was used to classify samples

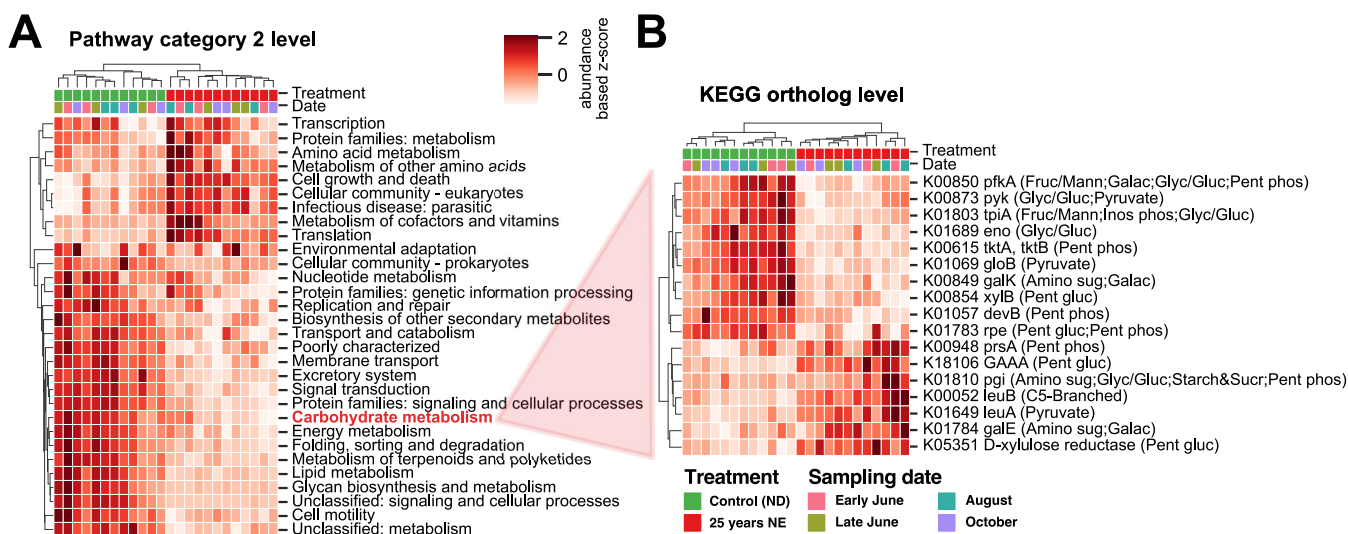


FIG 6 Functional analysis—random forest results. (A) Heat map of the 50% KEGG orthologs with the highest importance, summed to pathway category 2 level. Normalized expression values are converted to z-scores per row to highlight differences between samples. Hierarchical clustering of samples and KEGG pathways was performed using correlation metrics and complete linkage clustering. Colors on top indicate treatment and sampling date. (B) KEGG orthologs falling into the “Carbohydrate metabolism” pathway category. Visualization was performed as described for panel A.

by treatment and date. The 30 species having the highest importance in both data sets were then compared (Fig. 5B and C). For a more detailed description of the random forest results, see Text S1, Fig. S4 and S5, and Table S2. When classification of root samples by treatment type (ND, NE-5, and NE-25) was compared, predictive accuracy was high (>0.7) in both the ITS and RNA data. When only ND and NE-25 samples were used, the accuracy increased to 1 in both data sets. Root samples could not be accurately classified by sampling date, congruent with the earlier ordination-based statistical tests. In the ITS data, the 30 most important species had a summed importance of 0.69 and 17 species fell among the 12 most abundant families in the ITS data set (Fig. 5A). In addition, the mean relative abundance of a species and its feature importance had a Spearman rank correlation coefficient of 0.79. For the averaged RNA data, the top 30 most important species had summed importance of 0.4 (Fig. 5C), and in contrast to the ITS data set, only 9 of the top 30 important species belonged to the most abundant families (Fig. 5B); the Spearman correlation coefficient for mean relative abundance and importance was only 0.13.

Functional annotation of RNA data provides functional insight into fungal-community activity. As described above for taxonomic assignment, random forest analysis was applied using functional expression profiles of fungal transcripts to classify samples. Normalized expression values for ORFs assigned to the kingdom Fungi were summed to the KEGG ortholog (KO) level. To gain insight as to which functional categories were important for separating ND and NE-25 samples, the expression of KOs with a combined importance of 0.5 was summed to higher level functional categories in the KEGG pathway hierarchy (Fig. 6A). This revealed that the transcription, translation, and amino acid metabolism categories had higher z-expression values in NE-25 samples, while, e.g., carbohydrate metabolism, nucleotide and lipid metabolism, and signal transduction categories were more highly expressed in control samples.

Evidence in the literature suggests that N addition leads to a decrease in tree below-ground carbon allocation to ECM (29, 30), accompanied by a decrease in ectomycorrhizal growth (31). The random forest results were therefore used to explore the “carbohydrate metabolism” KEGG pathway category. In total, 17 KOs from this pathway category were among the most important 50% of KOs (Fig. 6B). Noteworthy orthologs with decreased abundance after long-term NE included several key players in the major glucose conversion pathway glycolysis (6-phosphofructokinase, pyruvate kinase, and triosephosphate isomerase), the glycolysis-parallel pentose phosphate pathway (transketolase and

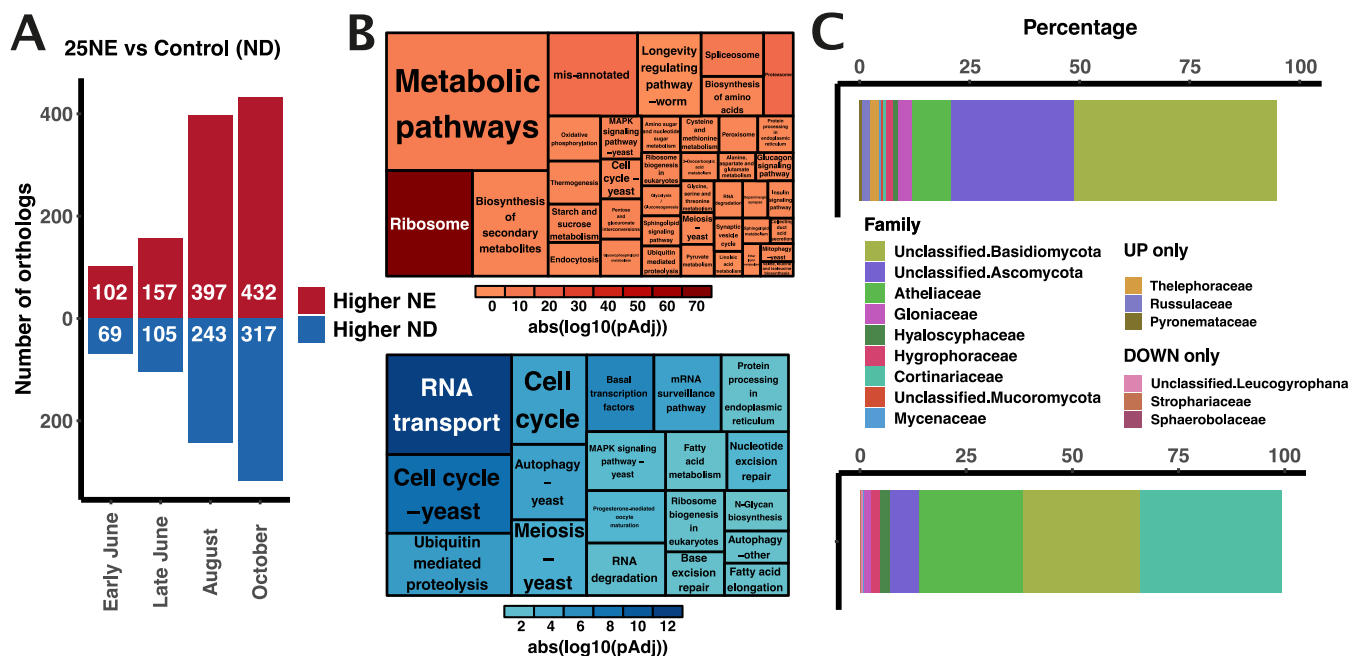


FIG 7 Functionally summarized differential KEGG ortholog (KO) abundance overview. (A) Bar plot showing the number of KOs found to be significantly more (red) or less (blue) abundant after 25 years of nutrient enrichment (NE) in comparison to the untreated controls. (B) KEGG pathway enrichment of all KOs identified as significantly differentially abundant at all four sampling dates, after 25 years of NE. Color intensity is determined by adjusted P value of enrichment, while rectangle size is proportional to the number of differentially abundant KOs in the respective pathway category. “Mis-annotated” refers to pathway categories associated with human diseases. Table S3 contains all terms and statistics. (C) Taxonomic distribution at the family level of all transcripts assigned to the set of KOs displayed in panel A. Only the 12 most abundant family annotations from both sets of KOs are displayed.

6-phosphogluconolactonase), and others from closely related downstream pathways. Some of the higher-abundance KOs were related to amino acid (specifically, leucine and isoleucine) metabolism: 3-isopropylaminate dehydrogenase and synthase.

Long-term nutrient enrichment leads to major functional changes at the community level with seasonal differences. A differential abundance analysis of the 25 years versus control condition identified KOs (1,822 KOs overall and 1,189 unique KOs), with an increasing number of differentially abundant KOs throughout the growing season (Fig. 7A). There were 47 commonly differentially abundant KOs at all four sampling dates (6 to 27% of KOs per seasonal time point), of which 29 increased and 18 decreased in abundance.

KEGG pathway enrichment of differentially abundant KOs (Fig. 7B) showed that general housekeeping pathways, including ribosome, proteasome, and spliceosome, were enriched among more highly abundant KOs, pointing to major shifts in biological activity. Pathways such as RNA transport, cell cycle, ubiquitin-mediated proteolysis, and mRNA surveillance were enriched among KOs with lower abundance. These enrichments were similar to results from the random forest analysis (Fig. 6). Notable and more specific pathways composed of KOs with higher abundance included starch and sucrose metabolism, biosynthesis of amino acids, and pentose and glucuronate interconversions. Specific pathways that were found to have significantly lower abundance after NE included autophagy, fatty acid elongation and metabolism, and N -glycan biosynthesis (Table S3).

Taxonomic annotations of the differentially abundant KOs provided an overview of which taxa were responsible for the observed functional changes (Fig. 7C). For instance, the family Atheliaceae, which did not seem to be strongly affected by NE in the ITS and RNA data sets (Fig. 5A), accounted for a much larger proportion (25% versus 10%) of the lower-abundance than the higher-abundance KOs.

Differential abundance analysis of the genus *Cortinarius* revealed extensive transcriptional downregulation. To highlight the ability to extract insight into the functional response of specific taxa, all transcripts assigned to the genus *Cortinarius* were selected. *Cortinarius* is a widespread and common genus of ECM fungi that has

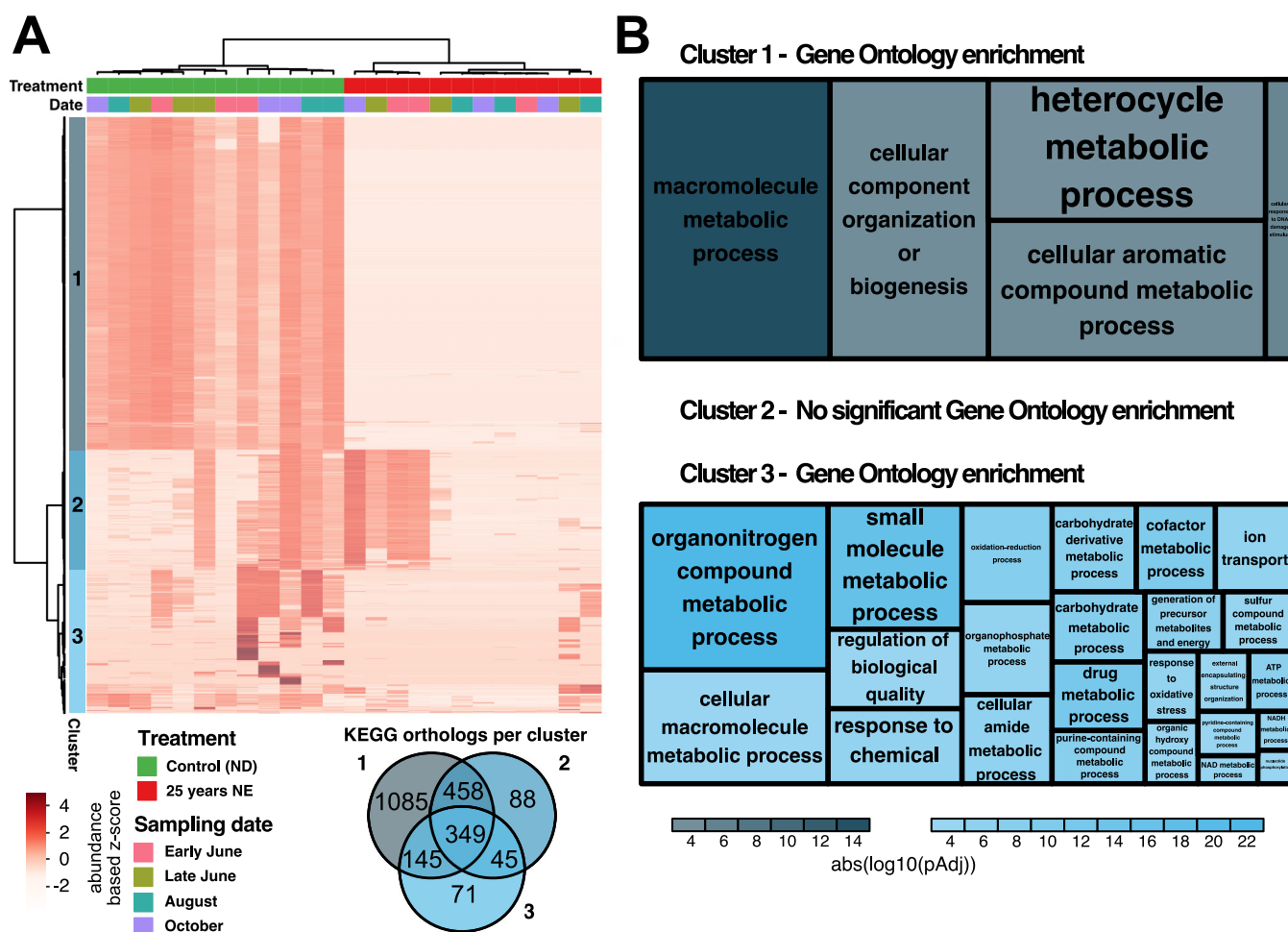


FIG 8 Hierarchical clustering of transcripts assigned to the genus *Cortinarius* and Gene Ontology (GO) enrichment. (A) Heat map of all transcripts assigned to the genus *Cortinarius*. Normalized variance stabilization transformed expression values were transformed to z-scores per row to highlight differences between samples. Hierarchical clustering of samples and transcripts was performed using Ward's minimum variance method. Colors on top indicate treatment and sampling date, while colors to the left indicate the three highest-level clusters. A color legend is provided below the heat map. The Venn diagram below the heat map shows KEGG orthologs (KOs) derived from the three clusters and the respective overlap between them. (B) Tree maps showing GO enrichment of all transcripts identified as belonging to the three highest level clusters. The upper tree map summarizes significant GO enrichments for cluster 1 (4,047 transcripts), and the lower tree map shows the enrichment for cluster 3 (1,741 transcripts). There were no significant enrichments for cluster 2 (1,465 transcripts). Color intensity is determined by adjusted P value of the enrichment, while rectangle size is proportional to the number of transcripts mapping to the respective GO term. Table S3 shows all terms and corresponding statistics.

been previously reported to be negatively affected by N enrichment in different contexts (28, 32, 33). In the current study, *Cortinarius* was still observed at low abundances after 25 years of treatment. We used a hierarchical clustering method (Ward minimum variance) and visualized the gene expression patterns of *Cortinarius* in our data set (Fig. 8A). This identified three main clusters of genes that had similar expression abundance patterns. Cluster 1 (4,047 transcripts) contained genes showing high abundance in the unperturbed control plots, which mostly dropped beyond detection after long term NE. Cluster 2 (1,456 transcripts) displayed high gene abundance in part of the ND and part of the NE samples, without any apparent pattern, while being mostly absent in other samples. Genes in cluster 3 (1,741 transcripts) displayed erratic abundance, with subsets of the cluster showing very high abundance in some samples, while other subsets displayed high abundances in other samples, with a tendency to be more abundant in control samples. Summarizing transcripts to KOs where possible, we observed substantial functional overlap between the three clusters (Fig. 8A, Venn diagram).

Gene Ontology (GO) enrichment of the three clusters identified significant enrichments for cluster 1 and cluster 3 (Fig. 8B). Cluster 1 showed significant ($P < 0.001$) enrichment in functions associated with the processing of organic macromolecules,

including aromatic and heterocyclic compounds. Cluster 3 was enriched for a greater variety of GO terms, including metabolism of both small molecules and macromolecules, metabolism of organonitrogen compounds and organophosphates, and carbohydrate and carbohydrate derivative metabolism. Both cluster 1 and cluster 3 showed a reduction in abundance after long-term NE, but this effect was much more pronounced in cluster 1.

DISCUSSION

RNA-Seq, in principle, enables studies of both the composition of active microbial communities and the biological functions being expressed by the constituent members. While several pipelines for the analysis of metatranscriptome data have been published (34–37), including evaluation of how the results from such data compare to those of amplicon- or whole-genome shotgun-based methods, these studies primarily focused on bacterial communities or relatively low-complexity species mixes or used the total RNA pool (21, 37, 38). Other studies have used both RNA-Seq and amplicon sequencing together to describe microbial communities from a taxonomic and functional perspective (18, 39). Here, we performed a study to ascertain how polyadenylated mRNA-Seq-based metatranscriptomics and DNA amplicon-based metagenomics results compare when profiling the complex root-associated and phyllospheric fungal communities associated with the boreal forest tree species Norway spruce (*Picea abies*) under natural and perturbed nutrient conditions. To facilitate the analysis of the metatranscriptomics data, we implemented a reproducible bioinformatic workflow to assemble fungal transcripts from the RNA-Seq data and subsequently annotate the assembled transcripts both functionally and taxonomically (Fig. 1). Creation of this custom workflow was necessitated by the lack of available tools for this specific case where our *a priori* criteria were the ability to (i) separate fungal and host reads and (ii) perform a *de novo* assembly of transcripts. We chose to assemble transcripts over direct read-based alignment to increase query sequence length and to maximize the number of aligned reads (Fig. S6) and because representative databases are lacking for our samples. Among previously published workflows, the IMP (Integrated Meta-omic Pipeline) workflow (37) supports host filtering and *de novo* assembly but was written with the human bacterial microbiome in mind and does not offer additional analyses such as differential expression or multiple sample comparisons (40). For the purpose of this study, we reanalyzed previously published ITS amplicon sequencing results from the same samples (28) to reflect new developments in sequence processing algorithms (41, 42).

Comparison of the control samples from roots and needles in the two data sets at the transcript and SOTU level revealed strikingly similar patterns of between-sample variation (Fig. 2), despite the proportionally higher number of needle SOTUs compared to the low number of fungal transcripts in needle samples. The low number of fungal transcripts obtained from the phyllospheric samples likely resulted from (i) the orders-of-magnitude-lower fungal load in the phyllospheric samples, leading to a much lower ratio of host to fungal nucleic acid in the extracts (43–45), and (ii) the higher phyllospheric fungal richness (Fig. 2) leading to sparser transcript count data in the needle samples. This low signal-to-noise ratio propagated through all further analyses of the needle samples, and for this reason we concentrated on the root data for in-depth analyses. This highlights the importance of relative fungal load for the application of this approach.

A further limitation of metatranscriptomics studies, especially where communities are composed of nonmodel systems, is the relatively low availability of sequenced reference fungal genomes. At the time the analyses were performed, there were 1,164 sequenced fungal genomes available at the JGI MycoCosm resource (46). This number is increasing steadily (1,681 at the time of writing) but is still far from being sufficiently comprehensive to capture a substantial portion of the fungal diversity present in most ecosystems. While the use of ITS amplicon data has its own limitations and

methodological issues, it could be expected to yield a more comprehensive catalogue of taxa present in the sample due to the more established database resource currently available. Comparing the fungal metatranscriptomics assembly to the ITS data from the same samples, the taxonomic overlap was found to be small at lower taxonomic levels (Fig. 3). While only a relatively low proportion of transcripts and SOTUs were assigned at the species level (42 and 39%, respectively), this small overlap is likely to also stem from the already low overlap between the UNITE and the MycoCosm databases at low taxonomic levels, with only 40% of species present in the MycoCosm resource being represented in the UNITE database (Fig. 3A). This issue currently limits RNA and ITS comparability on taxonomic terms, but this is likely to improve as fungal genomes become increasingly available and as the UNITE database increases the number of included ITS sequences.

Furthermore, the two methods showed a consistent variation in evenness (Fig. 4C), and both identified a consistent decrease in the proportion of reads belonging to the 12 most abundant family annotations in response to nutrient enrichment (Fig. 5A). For both data sets, this appeared to reflect higher diversity and loss of dominance of certain groups of fungi, as reported previously for the ITS data set (28). That some families strongly differ in abundance between the two data sets might result from methodological bias in the ITS amplicon data (9), but it was shown previously that while most DNA and RNA data correlate fairly well, total gene expression abundances of some groups deviate in their levels from what could be expected when looking only at genomic DNA abundance (47). Moreover, it has been found that even ITS amplicons obtained from DNA and RNA in soil fungal communities yielded very different taxonomic compositions (38). Finally, a characteristic of mixed-species RNA-Seq is that transcript abundance captures both expression and species abundance; i.e., a higher abundance could stem from either higher gene expression per nucleus or a higher nucleus count. Current sequencing library generation protocols do not allow these two factors to be separated, although future strategies will likely overcome this, for example, through use of long-read sequencing technologies, such as Pacific Bioscience or Oxford Nanopore, that do not require assembly or via the use of unique molecular identifiers and transcript assembly algorithms that utilize this information.

While taxonomic overlap was low between the metatranscriptomics and ITS amplicon sequencing data, the congruence between unsupervised ordination methods was high, not only when comparing control samples but also when comparing the root sample clustering by nutrient status (Fig. 4A and B). The phyllospheric results are discussed in Text S1. As an additional approach to assess the similarities and differences between the two data types, we performed random forest classifications of NE-25 and ND samples in both data sets. We found that in a direct comparison, the random forest classification performed better on the ITS data, especially for needle samples. Congruent with the previous statistical tests, the random forest classifier found a strong effect of NE and no seasonal effect on root fungal communities. Another interesting observation was the notably higher correlation between mean relative abundance and importance in the ITS compared to RNA data. This could again be an artifact of the low taxonomic annotation in the RNA data, but potentially it indicates that the metatranscriptomic data enables a higher resolution by containing both taxonomic and expression abundance. Several previous studies have shown the importance of low abundance community members in a functional context (38, 48–50).

We applied random forest and differential abundance analyses to demonstrate use of the RNA-Seq data to provide both functional and taxonomic insights into the root-associated fungal community of Norway spruce and how it is affected by NE. Random forest classification accuracy when using KO counts, in comparison to taxonomic annotations at the species level, proved to be similarly good when classifying by treatment and slightly better when classifying by sampling date. The slightly higher accuracy for sampling date in both roots and needles when using functional profiles indicates that over the course of the season, shifts in functions expressed by fungal species are more

pronounced, and yield higher signal strength, than the turnover of the species themselves. Both random forest and differential ortholog and transcript abundance analyses identified a number of functional categories with enrichment for transcripts having increased or decreased abundance as a result of the treatment, with highly congruent results from the two methods (Fig. 6 and 7). Furthermore, both methods were used to provide taxonomically resolved insights, identifying species and families that were important in explaining the separation either of the two treatment conditions (Fig. 5 and 7) or of transcripts from a specific family having significant changes in relative abundance (Fig. 7).

Finally, we demonstrated that we can pick one taxon of interest and investigate its specific transcriptomic response to the experimental conditions with the same methods, in this case, the known nitrophobic genus *Cortinarius* (Fig. 8). *Cortinarius* is one of the most species-rich ECM genera, with hundreds of species occurring in Sweden alone, and belongs to a group of ECM fungi that exhibit medium-distance fringe-type exploration and that have been shown in several studies to be sensitive to N addition (33). This N sensitivity has been hypothesized to be caused partly by their reliance on mobilizing organic N sources, using oxidative enzymes for degradation (32, 51). The high carbon cost of this foraging strategy would become disadvantageous with high inorganic N availability, due to both the decreased allocation of tree carbon below-ground (30, 31) and a decrease in the energetical efficiency of oxidative enzymes (52). More recent comparative genomic studies have shown that *Cortinarius glaucopus* (the only sequenced European *Cortinarius* species to date) has retained an unusually high number of genes for plant cell wall-degrading enzymes in its genome, compared to most other ECM fungi (53). While the majority of *Cortinarius* transcripts in our data set showed a strong and uniform reduction in abundance after long-term NE, groups of genes had more varied, limited, or no response to the treatment (Fig. 8).

The large extent of functional overlap between the three identified gene clusters could suggest that each cluster represents a different *Cortinarius* species (or group of species), each of which has a different level of sensitivity to N addition. We observed an enrichment of GO terms associated with metabolic processing of aromatic compounds in the cluster of genes that showed the strongest consistent decrease in abundance after long-term NE (cluster 1), potentially indicating a strong reliance on the degradation of phenolic compounds (e.g., lignin derivatives) for this species, in line with the aforementioned literature. Cluster 2 (equal expression representation in both conditions) did not yield significant GO enrichments, which could suggest that some *Cortinarius* species are not as negatively affected by high N content. The third cluster showed variable representation in control samples and similarly variable (but overall reduced) representation after long-term NE. The higher number of enriched GO terms, including many important metabolic processes, such as carbohydrate and organonitrogen utilization, potentially indicates that this group of transcripts is from a *Cortinarius* species that relies on other enzymatic mechanisms to obtain N and that are not as dependent on tree derived C as the species in cluster 1. While these interpretations are, admittedly, speculative, they serve to highlight the additional power provided by a metatranscriptomic approach for enabling functionally informed insights and hypothesis generation to direct subsequent studies.

In conclusion, we have demonstrated that RNA-Seq metatranscriptomics, under the prerequisite of a sufficient microbial load in the sample of interest, has the potential to bypass the inherent limitations of ITS amplicon sequencing, especially with further technology development and availability of more extensive databases of sequenced genomes. We have shown that in terms of alpha and beta diversity, comparable results can be obtained using only metatranscriptomic data, while ITS amplicon data are still currently needed to provide a more complete taxonomic profile of the fungal community. While we did not consider this in the current study, the data generated additionally capture transcriptome dynamics in the host tree, enabling a plethora of additional analyses apart from what we demonstrated

here. In conjunction with host tree expression, as well as other microbial communities, the presented data set and the approach in general hold great potential to yield insights into the dynamics of multispecies and multidomain gene expression and their interactions.

MATERIALS AND METHODS

Sample collection, nucleic acid extraction, and sequencing. Samples were collected during the growing season in 2012 and stored at -80°C . The ITS1 amplicon sequencing data used in this study were published previously (28) and reanalyzed for this study. RNA was extracted from the same spruce root and needle samples in early 2013 and used for RNA sequencing; the details are described in Text S1. RNA was successfully sequenced from 214 samples, 107 root and 107 needle samples. For ideal comparability to the ITS data, replicates from within one on-site block were pooled, resulting in 36 pooled root and needle samples.

Metatranscriptomic workflow. Preprocessing and analysis of metatranscriptomic data were implemented in a Snakemake workflow available on Bitbucket (54); we ensured complete, hardware-agnostic reproducibility through implementation in both docker and singularity containers. A detailed description of software and parameters used in the workflow is available in Text S1. Briefly, raw reads were trimmed and filtered using cutadapt and Trimmomatic (55, 56). Read quality scores before and after preprocessing were assessed with FastQC and MultiQC (57, 58). Fungal reads were selected from preprocessed reads aligned using bowtie2 against the JGI MycoCosm database (46, 59) and TaxMapper against its own database (60). Reads aligning to JGI MycoCosm were filtered for host reads by bowtie2 alignments against the Norway spruce reference genome obtained from PlantGenIE (61). We deduplicated fungal read pairs using FastUniq (62). Subsequently, fungal reads were assembled using Megahit, Trans-ABYSS, and Trinity (63–65). Open reading frames (ORFs) in the Megahit assembly were identified using GeneMarkS-T (66). FeatureCounts from the subread package (67) was used to count the reads aligning within ORFs with bowtie2. Raw counts were normalized to transcripts per million (TPM) (68). Translated protein sequences were annotated using eggno-mapper in conjunction with the eggno-g database (69, 70), to obtain KEGG ortholog (KO) and Gene Ontology (GO) annotations (71). For taxonomic annotation, we used a database comprising proteins constructed from JGI MycoCosm, TaxMapper, and the *Hygrophorus russula* MG78 genome with genes predicted using Augustus (72). *Hygrophorus russula* was included to account for the high abundance of *Hygrophorus* at the field site, observed both in the ITS data and *in situ* sporocarp assessments. Taxonomy was assigned to transcripts using contigtax (73), which uses rank-specific thresholds (74) to infer lowest common ancestors, based on DIAMOND BLASTX searches (75).

Amplicon sequence data pipeline. The code needed to run the preprocessing and analysis of the amplicon sequencing data is available on GitHub (76). A detailed description is available in Text S1. In short, raw reads were demultiplexed using deML (77), and primer sequences were removed using cutadapt (55) after pooling technical replicates, as described previously (28). The R package dada2 (42) was used to filter and denoise the reads, before dereplicating them into amplicon sequencing variants (ASVs), merging overlapping forward and reverse reads, and removing chimeric sequences. The ITS1 region was cut out from the ASVs using ITSx (78) and subsequently clustered into Swarm operational taxonomic units (SOTUs) using Swarm (41). Taxonomy was assigned using naive Bayesian classifier implemented in dada2, with the UNITE database as a reference (79).

Analyses and visualizations. All further analyses were performed using R (80), unless otherwise specified. Visualizations were plotted using ggplot2, unless otherwise specified (81). Venn diagrams in Fig. 2 were created using the R package VennDiagram (82), and Venn diagrams and correlations in Fig. 3 were created using jupyter and matplotlib (83, 84). Detailed parameter information can be found in the above git repositories, and a more detailed description is provided in Text S1. Amplicon sequencing data were filtered and rarefied using vegan (85), which was also used for PERMANOVA, Shannon diversity, and Mantel and Procrustes tests. The package phyloseq (v 1.28) was used to visualize PCoA ordinations (86). Linear mixed-effect models were used to test for significant differences in diversity using the nlme package (87) and the multcomp package (88). RNA-Seq-derived transcripts were selected to be of fungal origin and subsequently filtered using the same criteria as for the SOTUs. After filtering, the replicates per plot were merged by mean value to make the data more comparable to the ITS amplicon data. Filtered metatranscriptome count data were transformed using the function varianceStabilizingTransformation from the DESeq2 package prior to principal-component analysis (89). Random forest analyses were implemented using the RandomForestClassifier from scikit-learn (90). Heat maps summarizing random forest results were plotted using matplotlib (84), while the heat map in Fig. 8 was plotted using the R package pheatmap (91). DESeq2 was used to identify differentially abundant KOs and transcripts (89). Differentially abundant KOs and transcripts were filtered to have a log fold change of at least 0.5 and a P value of <0.05 . Functions for easier filtering and visualization of differential expression results were pulled from the Rtoolbox repository (92). The tool gofer2 (93) was used for KO and Gene Ontology enrichment, the R wrapper of which was pulled from the repository of the Umeå Plant Science Centre bioinformatics facility (94). The R package treemap was used to visualize the enrichments (95).

Data availability. The raw data from the ITS1 amplicon sequencing have been deposited in the European Nucleotide Archive (ENA) with accession number PRJEB21692 (96). RNA-Seq raw data are also

deposited in the ENA with accession number [PRJEB35783](https://www.ebi.ac.uk/ena/record/PRJEB35783) (97). Workflows and scripts to preprocess and analyze the data have been made available in the git repositories mentioned above.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

TEXT S1, DOCX file, 0.1 MB.

FIG S1, EPS file, 1.4 MB.

FIG S2, EPS file, 1 MB.

FIG S3, EPS file, 0.3 MB.

FIG S4, EPS file, 1.9 MB.

FIG S5, EPS file, 2.2 MB.

FIG S6, EPS file, 0.5 MB.

TABLE S1, XLSX file, 0.1 MB.

TABLE S2, XLSX file, 0.01 MB.

TABLE S3, XLSX file, 0.02 MB.

ACKNOWLEDGMENTS

We thank the Umeå Plant Science Centre Bioinformatics platform (UPSCb) for support and discussion. We also thank VINNOVA for funding of the Vinnova competence center, financial support of Andreas Schneider, and project operation funding. We thank the Swedish Government's strategic initiative "Trees and Crops for the Future." We acknowledge support from the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. John Sundh is financially supported by the Knut and Alice Wallenberg Foundation as part of the National Bioinformatics Infrastructure Sweden at Science for Life Laboratory.

We declare that we have no competing interests.

A.N.S.: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. J.S.: Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – review & editing. G.S.: Conceptualization, Formal analysis. K.R.: Investigation. N.D.: Resources, Software, Writing – review & editing. M.G.: Conceptualization, Software, Supervision. V.H.: Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review & editing. N.R.S.: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

REFERENCES

- Bacon CW, White JF. 2016. Functions, mechanisms and regulation of endophytic and epiphytic microbial communities of plants. *Symbiosis* 68:87–98. <https://doi.org/10.1007/s13199-015-0350-2>.
- López-Mondéjar R, Kostovčík M, Lladó S, Carro L, García-Fraile P. 2017. Exploring the plant microbiome through multi-omics approaches, p 233–268. In Kumar V, Kumar M, Sharma S, Prasad R (ed), *Probiotics in agroecosystem*. Springer, Singapore.
- Bálint M, Bahram M, Eren AM, Faust K, Fuhrman JA, Lindahl B, O'Hara RB, Öpik M, Sogin ML, Unterseher M, Tedersoo L. 2016. Millions of reads, thousands of taxa: microbial community structure and associations analyzed via marker genes. *FEMS Microbiol Rev* 40:686–700. <https://doi.org/10.1093/femsre/fuw017>.
- Huber JA, Mark Welch DB, Morrison HG, Huse SM, Neal PR, Butterfield DA, Sogin ML. 2007. Microbial population structures in the deep marine biosphere. *Science* 318:97–100. <https://doi.org/10.1126/science.1146689>.
- Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Bolchacova E, Voigt K, Crous PW, Miller AN, Wingfield MJ, Aime MC, An K-D, Bai F-Y, Barreto RW, Begerow D, Bergeron M-J, Blackwell M, Boekhout T, Bogale M, Boonyuen N, Burgaz AR, Buyck B, Cai L, Cai Q, Cardinali G, Chaverri P, Coppins BJ, Crespo A, Cubas P, Cummings C, Damm U, de Beer ZW, de Hoog GS, Del-Prado R, Dentinger B, Dieguez-Urbeondo J, Divakar PK, Douglas B, Duenas M, Duong TA, Eberhardt U, Edwards JE, Elshahed MS, Fliegerova K, Furtado M, Garcia MA, Ge Z-W, Griffith GW, Fungal Barcoding Consortium, et al. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc Natl Acad Sci U S A* 109:6241–6246. <https://doi.org/10.1073/pnas.1117018109>.
- Abarenkov K, Henrik Nilsson R, Larsson K-H, Alexander IJ, Eberhardt U, Erland S, Høiland K, Kjølner R, Larsson E, Pennanen T, Sen R, Taylor AFS, Tedersoo L, Ursing BM, Vrålstad T, Liimatainen K, Peintner U, Kõljalg U. 2010. The UNITE database for molecular identification of fungi—recent updates and future perspectives. *New Phytol* 186:281–285. <https://doi.org/10.1111/j.1469-8137.2009.03160.x>.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41: D590–D596. <https://doi.org/10.1093/nar/gks1219>.
- Acinas SG, Marcelino LA, Klepac-Ceraj V, Polz MF. 2004. Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrm operons. *J Bacteriol* 186:2629–2635. <https://doi.org/10.1128/JB.186.9.2629-2635.2004>.

9. Tedersoo L, Anslan S, Bahram M, Pölme S, Riit T, Liiv I, Kõljalg U, Kisand V, Nilsson H, Hildebrand F, Bork P, Abarenkov K. 2015. Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycKeys* 10:1–43. <https://doi.org/10.3897/mycokeys.10.4852>.
10. Carini P, Marsden PJ, Leff JW, Morgan EE, Strickland MS, Fierer N. 2016. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nat Microbiol* 2:16242. <https://doi.org/10.1038/nmicrobiol.2016.242>.
11. Delhomme N, Sundström G, Zamani N, Lantz H, Lin YC, Hvidsten TR, Höppner MP, Jern P, Van De Peer Y, Lundeberg J, Grabherr MG, Street NR. 2015. Serendipitous meta-transcriptomics: the fungal community of Norway spruce (*Picea abies*). *PLoS One* 10:e0139080. <https://doi.org/10.1371/journal.pone.0139080>.
12. Gonzalez E, Pitre FE, Pagé AP, Marleau J, Nissim WG, St-Arnaud M, Labrecque M, Joly S, Yergeau E, Brereton NJB. 2018. Trees, fungi and bacteria: tripartite metatranscriptomics of a root microbiome responding to soil contamination. *Microbiome* 6:1–30. <https://doi.org/10.1186/s40168-018-0432-5>.
13. Mateus ID, Masclaux FG, Aletti C, Rojas EC, Savary R, Dupuis C, Sanders IR. 2019. Dual RNA-seq reveals large-scale non-conserved genotype × genotype-specific genetic reprogramming and molecular crosstalk in the mycorrhizal symbiosis. *ISME J* 13:1226–1238. <https://doi.org/10.1038/s41396-018-0342-3>.
14. Luo B, Gu W, Zhong J, Wang Y, Zhang G. 2015. Revealing crosstalk of plant and fungi in the symbiotic roots of sewage-cleaning *Eichhornia crassipes* using direct de novo metatranscriptomic analysis. *Sci Rep* 5:15407. <https://doi.org/10.1038/srep15407>.
15. Žifčáková L, Větrovský T, Howe A, Baldrian P. 2016. Microbial activity in forest soil reflects the changes in ecosystem properties between summer and winter. *Environ Microbiol* 18:288–301. <https://doi.org/10.1111/1462-2920.13026>.
16. Hesse CN, Mueller RC, Vuysich M, Gallegos-Graves LV, Gleasner CD, Zak DR, Kuske CR. 2015. Forest floor community metatranscriptomes identify fungal and bacterial responses to N deposition in two maple forests. *Front Microbiol* 6:337. <https://doi.org/10.3389/fmicb.2015.00337>.
17. Bashiardes S, Zilberman-Schapiro G, Elinav E. 2016. Use of metatranscriptomics in microbiome research. *Bioinform Biol Insights* 10:BBI.S34610. <https://doi.org/10.4137/BBI.S34610>.
18. Crump BC, Wojahn JM, Tomas F, Mueller RS. 2018. Metatranscriptomics and amplicon sequencing reveal mutualisms in seagrass microbiomes. *Front Microbiol* 9:388. <https://doi.org/10.3389/fmicb.2018.00388>.
19. Kuske CR, Hesse CN, Challacombe JF, Cullen D, Herr JR, Mueller RC, Tsang A, Vilgalys R. 2015. Prospects and challenges for fungal metatranscriptomics of complex communities. *Fungal Ecol* 14:133–137. <https://doi.org/10.1016/j.funeco.2014.12.005>.
20. Pascual N, Loux V, Derozier S, Martin V, Debroas D, Malouf S, Humbert J-F, Leloup J. 2015. Technical challenges in metatranscriptomic studies applied to the bacterial communities of freshwater ecosystems. *Genetica* 143:157–167. <https://doi.org/10.1007/s10709-014-9783-4>.
21. Cottier F, Srinivasan KG, Yurieva M, Liao W, Poidinger M, Zolezzi F, Pavelka N. 2018. Advantages of meta-total RNA sequencing (MeTRS) over shotgun metagenomics and amplicon-based sequencing in the profiling of complex microbial communities. *NPJ Biofilms Microbiomes* 4:2–7. <https://doi.org/10.1038/s41522-017-0046-x>.
22. Tamm CO. 1991. Nitrogen in terrestrial ecosystems. Springer, Berlin, Germany.
23. Taylor AFS, Martin F, Read DJ. 2000. Fungal diversity in ectomycorrhizal communities of Norway spruce [*Picea abies* (L.) Karst.] and beech (*Fagus sylvatica* L.) along north-south transects in Europe, p 343–365. Springer, Berlin, Germany.
24. Read DJ, Leake JR, Perez-Moreno J. 2004. Mycorrhizal fungi as drivers of ecosystem processes in heathland and boreal forest biomes. *Can J Bot* 82:1243–1263. <https://doi.org/10.1139/b04-123>.
25. Grinhut T, Hadar Y, Chen Y. 2007. Degradation and transformation of humic substances by saprotrophic fungi: processes and mechanisms. *Fungal Biol Rev* 4:179–189.
26. Linder S. 1995. Foliar analysis for detecting and correcting nutrient imbalances in Norway spruce. *Ecol Bull* 44:178–190.
27. Bergh J, Linder S, Lundmark T, Elfving B. 1999. The effect of water and nutrient availability on the productivity of Norway spruce in northern and southern Sweden. *For Ecol Manage* 119:51–62. [https://doi.org/10.1016/S0378-1127\(98\)00509-X](https://doi.org/10.1016/S0378-1127(98)00509-X).
28. Haas JC, Street NR, Sjödin A, Lee NM, Högberg MN, Näsholm T, Hurry V. 2018. Microbial community response to growing season and plant nutrient optimisation in a boreal Norway spruce forest. *Soil Biol Biochem* 125:197–209. <https://doi.org/10.1016/j.soilbio.2018.07.005>.
29. Litton CM, Raich JW, Ryan MG. 2007. Carbon allocation in forest ecosystems. *Global Change Biol* 13:2089–2109. <https://doi.org/10.1111/j.1365-2486.2007.01420.x>.
30. Iivonen S, Kaakinen S, Jolkonen A, Vapaavuori E, Linder S. 2006. Influence of long-term nutrient optimization on biomass, carbon, and nitrogen acquisition and allocation in Norway spruce. *Can J Res* 36:1563–1571. <https://doi.org/10.1139/x06-035>.
31. Treseder KK. 2004. A meta-analysis of mycorrhizal responses to nitrogen, phosphorus, and atmospheric CO₂ in field studies. *New Phytol* 164:347–355. <https://doi.org/10.1111/j.1469-8137.2004.01159.x>.
32. Bödeker ITM, Clemmensen KE, de Boer W, Martin F, Olson Å, Lindahl BD. 2014. Ectomycorrhizal *Cortinarius* species participate in enzymatic oxidation of humus in northern forest ecosystems. *New Phytol* 203:245–256. <https://doi.org/10.1111/nph.12791>.
33. Lilleskov EA, Hobbie EA, Horton TR. 2011. Conservation of ectomycorrhizal fungi: exploring the linkages between functional and taxonomic responses to anthropogenic N deposition. *Fungal Ecol* 4:174–183. <https://doi.org/10.1016/j.funeco.2010.09.008>.
34. Martinez X, Pozuelo M, Pascal V, Campos D, Gut I, Gut M, Azpiroz F, Guarnier F, Manichanh C. 2016. MetaTrans: an open-source pipeline for metatranscriptomics. *Sci Rep* 6:26447. <https://doi.org/10.1038/srep26447>.
35. Halliday S, Parkinson J. 2016. Gist: an ensemble approach to the taxonomic classification of metatranscriptomic sequence data. *bioRxiv* e081026. <https://doi.org/10.1101/081026>.
36. Westreich ST, Treiber ML, Mills DA, Korf I, Lemay DG. 2017. SAMSA2: a standalone metatranscriptome analysis pipeline. *BMC Bioinformatics* 19:175. <https://doi.org/10.1101/195826>.
37. Narayanasamy S, Jarosz Y, Muller EEL, Heintz-Buschart A, Herold M, Kaysen A, Laczny CC, Piel N, May P, Wilmes P. 2016. IMP: a pipeline for reproducible reference-independent integrated metagenomic and metatranscriptomic analyses. *Genome Biol* 17:260. <https://doi.org/10.1186/s13059-016-1116-8>.
38. Baldrian P, Kolařík M, Stursová M, Kopecký J, Valášková V, Větrovský T, Žifčáková L, Snajdr J, Řídl J, Vlček C, Voříšková J. 2012. Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *ISME J* 6:248–258. <https://doi.org/10.1038/ismej.2011.95>.
39. Song Z, Du H, Zhang Y, Xu Y. 2017. Unraveling core functional microbiota in traditional solid-state fermentation by high-throughput amplicons and metatranscriptomics sequencing. *Front Microbiol* 8:1294. <https://doi.org/10.3389/fmicb.2017.01294>.
40. Shakya M, Lo CC, Chain PSG. 2019. Advances and challenges in metatranscriptomic analysis. *Front Genet* 10:904. <https://doi.org/10.3389/fgene.2019.00904>.
41. Mahé F, Rognes T, Quince C, De Vargas C, Dunthorn M. 2015. Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ* 3:e1420. <https://doi.org/10.7717/peerj.1420>.
42. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581–583. <https://doi.org/10.1038/nmeth.3869>.
43. Bulgarelli D, Rott M, Schlaeppi K, Ver Loren van Themaat E, Ahmadinejad N, Assenza F, Rauf P, Huettel B, Reinhardt R, Schmelzer E, Peplies J, Gloeckner FO, Amann R, Eickhorst T, Schulze-Lefert P. 2012. Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature* 488:91–95. <https://doi.org/10.1038/nature11336>.
44. Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, Tremblay J, Engelbrekton A, Kunin V, del Rio TG, Edgar RC, Eickhorst T, Ley RE, Hugenholtz P, Tringe SG, Dangl JL. 2012. Defining the core Arabidopsis thaliana root microbiome. *Nature* 488:86–90. <https://doi.org/10.1038/nature11237>.
45. Sakai M, Ikenaga M. 2013. Application of peptide nucleic acid (PNA)-PCR clamping technique to investigate the community structures of rhizobacteria associated with plant roots. *J Microbiol Methods* 92:281–288. <https://doi.org/10.1016/j.mimet.2012.09.036>.
46. Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otilar R, Riley R, Salamov A, Zhao X, Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I. 2014. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res* 42:D699–D704. <https://doi.org/10.1093/nar/gkt1183>.
47. Schirmer M, Franzosa EA, Lloyd-Price J, McIver LJ, Schwager R, Poon TW, Ananthakrishnan AN, Andrews E, Barron G, Lake K, Prasad M, Sauk J,

- Stevens B, Wilson RG, Braun J, Denson LA, Kugathasan S, McGovern DPB, Vlamakis H, Xavier RJ, Huttenhower C. 2018. Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat Microbiol* 3:337–346. <https://doi.org/10.1038/s41564-017-0089-z>.
48. Benjamino J, Lincoln S, Srivastava R, Graf J. 2018. Low-abundant bacteria drive compositional changes in the gut microbiota after dietary alteration. *Microbiome* 6:86. <https://doi.org/10.1186/s40168-018-0469-5>.
49. Dawson W, Hör J, Egert M, van Kleunen M, Pester M. 2017. A small number of low-abundance bacteria dominate plant species-specific responses during rhizosphere colonization. *Front Microbiol* 8:975. <https://doi.org/10.3389/fmicb.2017.00975>.
50. Hausmann B, Knorr K-H, Schreck K, Tringe SG, Glavina del Rio T, Loy A, Pester M. 2016. Consortia of low-abundance bacteria drive sulfate reduction-dependent degradation of fermentation products in peat soil microcosms. *ISME J* 10:2365–2375. <https://doi.org/10.1038/ismej.2016.42>.
51. Kyaschenko J, Clemmensen KE, Hagenbo A, Karlton E, Lindahl BD. 2017. Shift in fungal communities and associated enzyme activities along an age gradient of managed *Pinus sylvestris* stands. *ISME J* 11:863–874. <https://doi.org/10.1038/ismej.2016.184>.
52. Bonner MT, Castro D, Schneider AN, Sundström G, Hurry V, Street NR, Näsholm T. 2019. Why does nitrogen addition to forest soils inhibit decomposition? *Soil Biol Biochem* 137:107570. <https://doi.org/10.1016/j.soilbio.2019.107570>.
53. Miyauchi S, Kiss E, Kuo A, Drula E, Kohler A, Sánchez-García M, Morin E, Andreopoulos B, Barry KW, Bonito G, Buée M, Carver A, Chen C, Cichocki N, Clum A, Culley D, Crous PW, Fauchery L, Girlanda M, Hayes RD, Kéri Z, LaButti K, Lipzen A, Lombard V, Magnuson J, Maillard F, Murat C, Nolan M, Ohm RA, Pangilinan J, de Pereira MF, Perotto S, Peter M, Pfister S, Riley R, Sitrif Y, Stielow JB, Szöllösi G, Žifčáková L, Štursová M, Spatafora JW, Tedersoo L, Vaario LM, Yamada A, Yan M, Wang P, Xu J, Bruns T, Baldrian P, Vilgalys R, Dunand C, Henrissat B, Grigoriev IV, Hibbett D, et al. 2020. Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. *Nat Commun* 11:5125. <https://doi.org/10.1038/s41467-020-18795-w>.
54. Sundh J. 2020. N_Street_1801. Bitbucket. https://bitbucket.org/scilifelab-lts/n_street_1801/src/master/. Accessed 25 June 2020.
55. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet J* 17:10. <https://doi.org/10.14806/ej.17.1.200>.
56. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
57. Andrews S, Krueger F, Secondos-Pichon A, Biggins F, Wingett S. 2015. FastQC. A quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute.
58. Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32:3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>.
59. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
60. Beisser D, Graupner N, Grossmann L, Timm H, Boenigk J, Rahmann S. 2017. TaxMapper: an analysis tool, reference database and workflow for metatranscriptome analysis of eukaryotic microorganisms. *BMC Genomics* 18:787. <https://doi.org/10.1186/s12864-017-4168-6>.
61. Sundell D, Mannapperuma C, Netotea S, Delhomme N, Lin Y-C, Sjödin A, Van de Peer Y, Jansson S, Hvidsten TR, Street NR. 2015. The Plant Genome Integrative Explorer Resource: PlantGenIE.org. *New Phytol* 208:1149–1156. <https://doi.org/10.1111/nph.13557>.
62. Xu H, Luo X, Qian J, Pang X, Song J, Qian G, Chen J, Chen S. 2012. FastUniq: a fast de novo duplicates removal tool for paired short reads. *PLoS One* 7:e52249. <https://doi.org/10.1371/journal.pone.0052249>.
63. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>.
64. Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, Mungall K, Lee S, Okada HM, Qian JQ, Griffith M, Raymond A, Thiessen N, Cezard T, Butterfield YS, Newsome R, Chan SK, She R, Varhol R, Kamoh B, Prabhu A-L, Tam A, Zhao Y, Moore RA, Hirst M, Marra MA, Jones SJM, Hoodless PA, Birol I. 2010. De novo assembly and analysis of RNA-seq data. *Nat Methods* 7:909–912. <https://doi.org/10.1038/nmeth.1517>.
65. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652. <https://doi.org/10.1038/nbt.1883>.
66. Tang S, Lomsadze A, Borodovsky M. 2015. Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res* 43:e78. <https://doi.org/10.1093/nar/gkv227>.
67. Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30:923–930. <https://doi.org/10.1093/bioinformatics/btt656>.
68. Wagner GP, Kin K, Lynch VJ. 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* 131:281–285. <https://doi.org/10.1007/s12064-012-0162-3>.
69. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol* 34:2115–2122. <https://doi.org/10.1093/molbev/msx148>.
70. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 47:D309–D314. <https://doi.org/10.1093/nar/gky1085>.
71. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44:D457–D462. <https://doi.org/10.1093/nar/gkv1070>.
72. Stanke N, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* 34:W435–W439. <https://doi.org/10.1093/nar/gkl200>.
73. Sundh J. 2020. contigtax. Github. <https://github.com/NBISweden/contigtax>. Accessed 29 April 2020.
74. Luo C, Rodriguez-R LM, Konstantinidis KT. 2014. MyTaxa: an advanced taxonomic classifier for genomic and metagenomic sequences. *Nucleic Acids Res* 42:e73. <https://doi.org/10.1093/nar/gku169>.
75. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12:59–60. <https://doi.org/10.1038/nmeth.3176>.
76. Schneider AN. 2020. its_workflow. Github. https://github.com/andnischneider/its_workflow. Accessed 15 December 2020.
77. Renaud G, Stenzel U, Maricic T, Wiebe V, Kelso J. 2015. deML: robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics* 31:770–772. <https://doi.org/10.1093/bioinformatics/btu719>.
78. Bengtsson-Palme J, Ryberg M, Hartmann M, Branco S, Wang Z, Godhe A, De Wit P, Sánchez-García M, Ebersberger I, de Sousa F, Amend AS, Jumpponen A, Unterseher M, Kristiansson E, Abarenkov K, Bertrand YJK, Sanli K, Eriksson KM, Vik U, Veldre V, Nilsson RH. 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 Ecol Evol* 4:914–919. <https://doi.org/10.1111/2041-210X.12073>.
79. Kõljalg U, Larsson KH, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, Erland S, Høiland K, Kjoller R, Larsson E, Pennanen T, Sen R, Taylor AFS, Tedersoo L, Vrålstad T, Ursing BM. 2005. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytol* 166:1063–1068. <https://doi.org/10.1111/j.1469-8137.2005.01376.x>.
80. R Core Team. 2019. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
81. Wickham H. 2009. ggplot2: elegant graphics for data analysis. Springer-Verlag, New York, NY.
82. Chen H, Boutros PC. 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* 12:35. <https://doi.org/10.1186/1471-2105-12-35>.
83. Kluver T, Ragan-Kelley B, Pérez F, Granger B, Bussonnier M, Frederic J, Kelley K, Hamrick J, Grout J, Corlay S, Ivanov P, Avila D, Abdalla S, Willing C, Development Team J. 2016. Jupyter Notebooks—a publishing format for reproducible computational workflows, p 87–90. In Loizides F, Schmidt B (ed), Positioning and power in academic publishing: players, agents and agendas. IOS Press, Amsterdam, The Netherlands.
84. Hunter JD. 2007. Matplotlib: a 2D graphics environment. *Comput Sci Eng* 9:99–104.
85. Dixon P. 2003. VEGAN, a package of R functions for community ecology. *J Veg Sci* 14:927–930. <https://doi.org/10.1111/j.1654-1103.2003.tb02228.x>.
86. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. <https://doi.org/10.1371/journal.pone.0061217>.

87. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. 2020. nlme: linear and nonlinear mixed effects models.
88. Hothorn T, Bretz F, Westfall P. 2008. Simultaneous inference in general parametric models. *Biom J* 50:346–363. <https://doi.org/10.1002/bimj.200810425>.
89. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550. <https://doi.org/10.1186/s13059-014-0550-8>.
90. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. 2011. Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830.
91. Kolde R. 2012. Pheatmap: pretty heatmaps. R Package 1
92. Serrano A. 2020. R toolbox. Github. <https://github.com/loalon/Rtoolbox>. Accessed 15 August 2020.
93. Schiffthaler B. 2018. Gofer 2. Github. <https://github.com/bschiffthaler/gofer2>. Accessed 25 June 2020.
94. UPSCb. 2020. UPSCb-common. Github. <https://github.com/UPSCb/UPSCb-common>. Accessed 15 August 2020.
95. Tennekes M, Ellis P. 2017. Package 'treemap.' R.
96. Haas JC. 2018. Study of fungal and prokaryotic communities in a long-term nutrient optimised boreal spruce forest. ENA accession number PRJEB21692. <https://www.ebi.ac.uk/ena/browser/view/PRJEB21692>.
97. Schneider AN. 2020. Seasonal root and needle transcriptomics 2012. ENA accession number PRJEB35783. <https://www.ebi.ac.uk/ena/browser/view/PRJEB35783>.